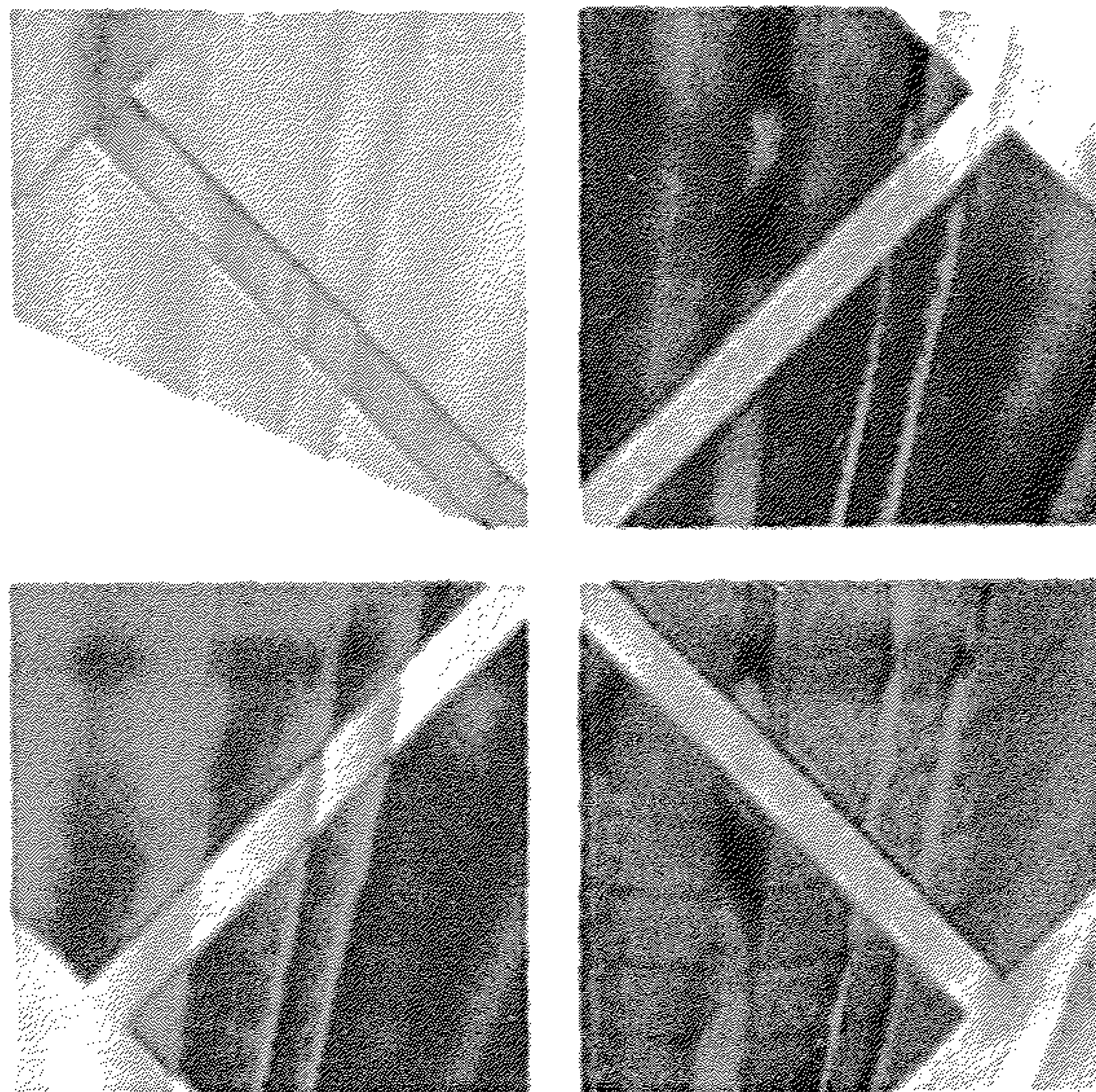
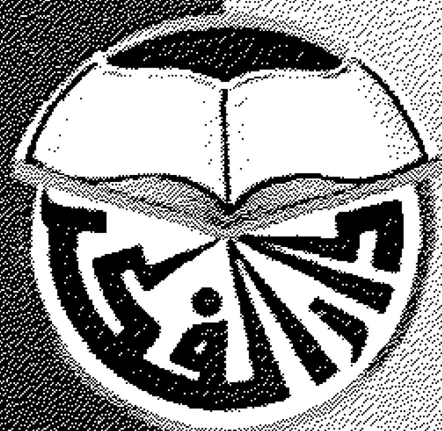
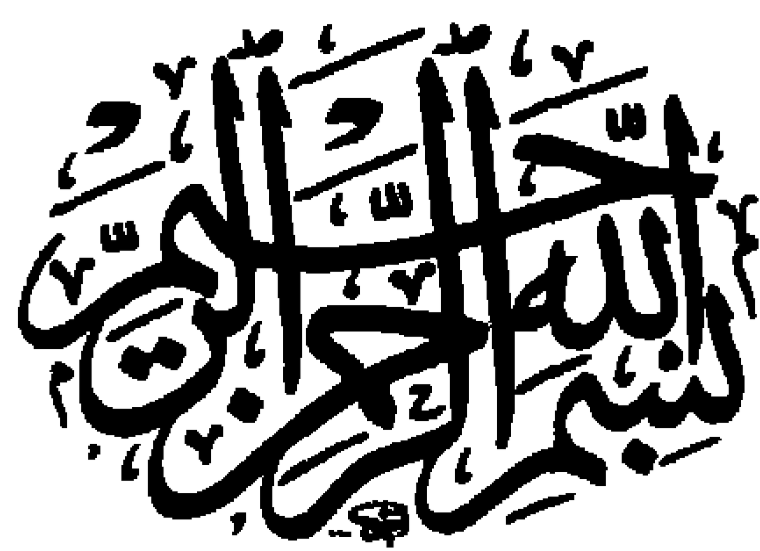
الإختبارات والمقاييس التربوية والنفسية



الدكتور
صلاح الدين محمود علام





الاختبارات والمقاييس التربويـــة والنفسيـة

رقم التصنيف: 370.15

رقم الايداع لدى دائرة المكتبة الوطنية: 2005/9/2167

المؤلف ومن هو في حكمه: صلاح الدين محمود علام

عنـــوان الكتـــــــــاب: الاختبارات والمقاييس: النفسية والتربوية

الموضـــــــــوع الرئيسي: علم النفس التربوي// الاختبارات

// التربية // علم النفس

بيانـــــــــات النشـــر: عمان - دار الفكـــر للنشـــر والتوزيـــع

* تم اعداد بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية

ISBN 9957-07- 437-7

الطبعة الأولى

1426 - 2006



سوق البتراء (الحجيري) ساحة الجامع الحسيني

هاتف: 4621938 - فاكس: 4654761

ص.ب : 183520 عمان 11118 الأردن

Email: info@daralfiker.com

المطابع المركزية - الأردن

الإختبارات والمقاييس التربوية والنفسية

تأليف

الاستاذ الدكتور صلاح الدين محمود علام

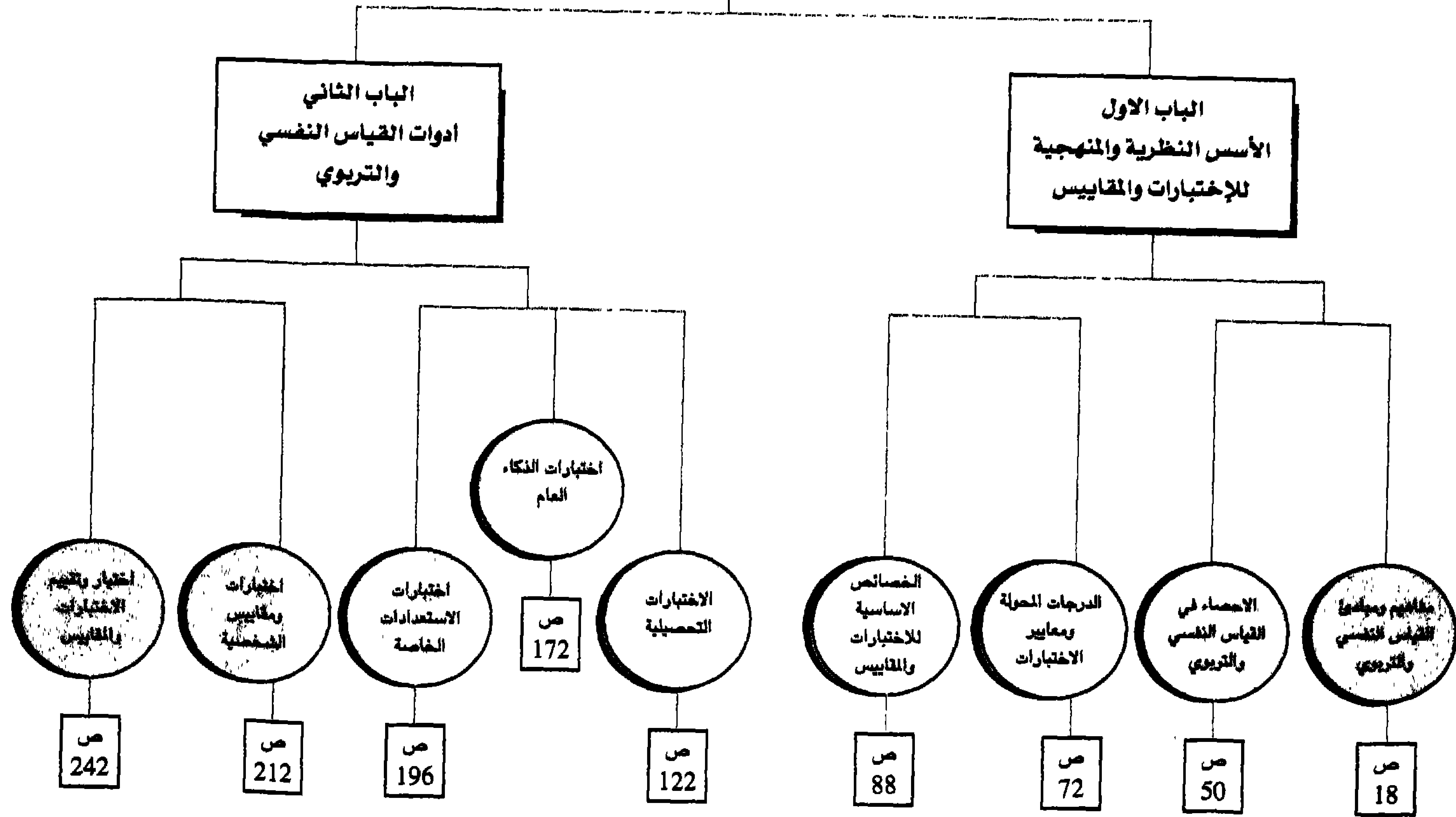
أستاذ القياس والتقويم والاحصاء التربوي والنفسي

كلية التربية - جامعة الازهر

الطبعة الأولى

1426-2006

دار الفكر
ناشرون وموزعون


ابواب وفصول الكتاب
الباب الاول
الأسس النظرية والمنهجية
للإختبارات والمقاييس
الباب الثاني
ادوات القياس النفسي
والتربوي
مفاهيم ومبادئ القياس النفسي والتربوي
ص 18
الاحصاء في القياس النفسي والتربوي
ص 50
الدرجات المحولة ومعايير الاختبارات
ص 72
الخصائص الاساسية للاختبارات والمقاييس
ص 88
الاختبارات التحصيلية
ص 122
اختبارات الذكاء العام
ص 172
اختبارات الاستعدادات الخاصة
ص 196
اختبارات ومقاييس الشخصية
ص 212
اختيار وتقييم الاختبارات والمقاييس
ص 242

الفهرس

مقدمة

الباب الأول: الأسس النظرية والمنهجية للاختبارات والمقاييس

الفصل الأول: مفاهيم ومبادىء القياس النفسي والتربوي

تقديم

يعد قياس السمات والخصائص الإنسانية من المجالات الرئيسة في العلوم النفسية والتربوية. فالأخصائيون النفسيون، والمربون، والمعنيون بانتقاء الأفراد وتوجيههم وتسكينهم في مختلف ميادين العمل والأنشطة المهنية والفنية يهتمون بقياس وتقييم الجوانب المعرفية والشخصية وغيرها لدى هؤلاء الأفراد. ويستخدمون في ذلك اختبارات ومقاييس متعددة ومتنوعة يزخر بها علم القياس النفسي والتربوي.

وقد أسهمت التطورات المتسارعة في منهجيات وتقنيات هذا العلم في تحديث طرق وأساليب بناء مختلف أنواع أدوات القياس، بما يؤدي إلى زيادة دقة وموضوعية قياس وتقييم السمات الانسانية المتعددة. وترتب على ذلك مزيداً من انتشار استخدام الاختبارات والمقاييس التربوية والنفسية في الأغراض المختلفة في كثير من الدول، وبخاصة الدول المتطورة. وأصبحت هذه الأدوات وما يتعلق بها من استخدامات مجالاً خصباً للدراسات والبحوث في مؤسسات اختبارية عالمية وإقليمية متخصصة وتُفرد لها حوليات موسوعية، كما أصبحت تنال قدراً وفيراً من الجدل والنقد من جانب علماء النفس والتربية، والفلاسفة، ورجال السياسة، بل ومن الجمهور بعامة.

ويتعلق ذلك بدرجة أساسية بمنطق قياس السمات والخصائص الإنسانية، وكيفية اختلافها عن قياس الخصائص الفيزيائية من حيث: طبيعة عملية القياس ومستوياته ومدى تكافؤ وحداته، وتأثير الاختبارات والمقاييس وما تستند إليها من أساليب وإجراءات في حياة الأفراد ومستقبلهم كأعضاء في مجتمع، وما قد يترتب على ذلك من إساءة استخدام المعلومات المستمدة منها ومحدوديتها.

ويُعد تصميم وبناء الاختبارات والمقاييس التربوية والنفسية علماً وفنّاً يتطلب اكتساب المشتغلين بالقياس، والمربين، وطلاب كليات التربية، والباحثين، والقائمين بتوظيف وإدارة الموارد البشرية، المعارف والمهارات المنهجية والمنطقية والإحصائية والتقنية الضرورية لاستيعاب المفاهيم والمبادىء الأساسية لعلم القياس النفسي والتربوي. كما أن الاستخدامات المناسبة لهذه الأدوات، والتوصل إلى استنتاجات صائبة يمكن الاستناد إليها بدرجة عالية من الثقة في اتخاذ قرارات تتعلق بالأفراد يتطلب تدريباً مكثفاً، وخبرة واسعة، وألفة بمصادر الاختبارات والمقاييس المتوافرة وأنواعها، واستخداماتها، ومحكات تقييم جودتها، وكيفية اختيار المناسب منها لغرض أو أغراض معينة.

لذلك فإن هذا الكتاب يعد مرجعاً أساسياً للطلاب، والمربين، والباحثين في علم النفس والتربية، والممارسين في مجال الاختبارات والمقاييس التربوية والنفسية.

ويتضمن الكتاب بابين رئيسين: يتناول الأول الأسس النظرية والمنهجية للاختبارات والمقاييس، ويتناول الثاني أدوات القياس التربوي والنفسي.

ويشتمل الباب الأول على أربعة فصول، يتناول الفصل الأول مفاهيم ومبادىء القياس النفسي والتربوي، من حيث: مفهوم القياس وأهميته، ومنطق السمات الإنسانية في القياس، وطبيعة القياس النفسي والتربوي، وخصائص موازين القياس، ومستويات أو موازين القياس، ومفهوم الاختبار كأداة قياس، وأنواع الاختبارات والمقاييس التربوية والنفسية، وخصائص عامة للاختبارات والمقاييس، ومفهوم التقويم وعلاقته بالقياس والاختبارات، وأخلاقيات استخدام الاختبارات.

ويتناول الفصل الثاني الإحصاء في القياس النفسي والتربوي متضمناً: أنواع البيانات الاختبارية، وتنظيم الدرجات، ومقاييس النزعة المركزية والعلاقة بينها، ومقاييس التشتت، والتوزيع الاعتدالي للدرجات، ومقاييس العلاقة، والارتباط والتنبؤ.

ويتناول الفصل الثالث الدرجات المحولة، ومعايير الاختبارات، من حيث: أنواع المعايير، والدرجات المعيارية، والمئينيات والإعشاريات، والتساعيات المعيارية، ونسب الذكاء، ومعايير الفرق الدراسية، والصفحات النفسية (البروفيلات).

ويتناول الفصل الرابع الخصائص الأساسية للاختبارات والمقاييس، متضمناً: مفهوم الثبات وأهميته، ومصادر أخطاء القياس التي تؤثر في الثبات، وطرق تقدير ثبات درجات الاختبارات، والخطأ المعياري للقياس، وبعض العوامل التي تؤثر في تفسير قيم معامل الثبات، ومفهوم الصدق وأهميته، وأنواع صدق الاختبار، وتعقيب عام على أنواع الصدق، وتحليل مفردات الاختبار وأهميته، ومعامل صعوبة وتمييز المفردة وتفسير ذلك.

ويشتمل الباب الثاني على اربعة فصول؛ يتناول الفصل الخامس الاختبارات التحصيلية من حيث: مفهوم الاختبار التحصيلي، والفروض الأساسية التي يرتكن إليها قياس التحصيل، وتصنيف الاختبارات التحصيلية، ومراحل بنائها، والاختبارات التحصيلية المقننة، وخصائصها، وبنائها، والفروق بينها وبين الاختبارات الصفية.

ويتناول الفصل السادس اختبارات الذكاء العام من حيث: مفهوم الذكاء وتطوراته التاريخية، والتعريفات المتعددة للذكاء، ونظريات بنية الذكاء، وتصنيف اختبارات الذكاء العام الفردية والجماعية، وتعقيب على هذه الاختبارات.

ويتناول الفصل السابع اختبارات الاستعدادات الخاصة، من حيث: أنواعها، واستخداماتها، وبطاريات الاستعدادات الخاصة المتعددة، واختبارات استعدادات وقدرات نوعية، تتضمن اختبارات الاستعدادات الميكانيكية، واختبارات الاستعداد الكتابي، واختبارات الاستعدادات الموسيقية والفنية.

ويتناول الفصل الثامن اختبارات ومقاييس الشخصية، من حيث: مفهوم

الشخصية، وأهداف قياسها، وبعض مشكلات القياس، واستراتيچيات بناء استبيانات الشخصية، وتعقيب عام على هذه الاستبيانات، والمقاييس الإسقاطية للشخصية، وتعقيب عام على هذه المقاييس.

ويتناول الفصل التاسع اختيار وتقييم الاختبارات والمقاييس متضمناً: اعتبارات عملية وفنية عند اختيارها، ودليل الاختبار، ومصادر الحصول على معلومات عن الاختبارات والمقاييس المنشورة.

وقد قدمت العديد من الأشكال التخطيطية التي توضح المفاهيم والمبادىء والأساليب المتعلقة بالموضوعات التي تناولها للاسترشاد بها، والاستزادة منها.

وأرجو من الله سبحانه وتعالى أن يفيد بهذا الكتاب طلاب كليات التربية، والباحثين، والمربين، ومستخدمي الاختبارات والمقاييس التربوية والنفسية في الأغراض الحيوية المختلفة.

ونسأل الله التوفيق والسداد

د. صلاح الدين محمود علام

القاهرة في أكتوبر 2005 م

الباب الأول 1

الأسس النظرية والمنهجية للاختبارات والمقاييس

- الفصل الأول: مفاهيم ومبادىء القياس النفسي والتربوي
- الفصل الثاني: الإحصاء في القياس النفسي والتربوي
- الفصل الثالث: الدرجات المحولة، ومعايير الاختبارات
- الفصل الرابع: الخصائص الأساسية للاختبارات والمقاييس

الفصل الأول

مفاهيم ومبادئ القياس النفسي والتربوي

- مقدمة
- مفهوم القياس وأهميته
- منطق السمات الإنسانية في القياس
- طبيعة القياس النفسي والتربوي
- خصائص موازين القياس
- مستويات أو موازين القياس
- مفهوم الاختبار كأداة قياس
- أنواع الاختبارات والمقاييس التربوية والنفسية
- خصائص عامة للاختبارات والمقاييس التربوية والنفسية
- استخدامات الاختبارات والمقاييس التربوية والنفسية
- مفهوم التقويم وعلاقته بالقياس والاختبارات
- أخلاقيات استخدام الاختبارات

مقدمة:

تتميز الدراسة العلمية الموضوعية للسلوك الإنساني باعتمادها اعتماداً أساسياً على الأساليب الكمية في قياس هذا السلوك بمختلف جوانبه. فبدون هذه الأساليب تقتصر دراسة السلوك الإنساني على الوصف اللفظي، والملاحظة الذاتية، والتأملات الفردية. فقد نشأ علم النفس مرتبطاً بالفلسفة وكان يُعد فرعاً من فروعها، ولكن عندما استُخدم التجريب والقياس في القرن التاسع عشر اعتماداً على أسس الطريقة العلمية بدلاً من الطرق الفلسفية كأسلوب للتفكير والبحث في الظواهر النفسية تطور علم النفس تطوراً كبيراً، حيث انتقل من دراسة مفاهيم عارضة مثل: الروح والنفس والعقل، إلى دراسة السلوك، وأصبح علماً مستقلاً عن الفلسفة.

وقد كان هذا التحول ضرورياً لأن السلوك الإنساني هو الشيء الأساسي الذي يمكن ملاحظته وتسجيله وقياسه، ومن ثم يمكن بحثه ودراسته. فالدراسة العلمية للسلوك الإنساني تعتمد على الملاحظة، والتجريب، والاستقصاء، للتوصل إلى القوانين والمبادىء العامة التي تساعد في وصف الظواهر السلوكية، وتفسيرها، والتنبؤ بها والتحكم فيها. كما أن علم النفس أصبح علماً عندما استند إلى الطريقة العلمية في دراسة الفروق بين الأفراد في مختلف السمات الإنسانية وتكميم هذه الفروق، أي قياسها قياساً كمياً، والتنبؤ بها.

ومما لا شك فيه أن العلوم السلوكية، وبخاصة علم النفس، لم تصل بعد إلى مرحلة الدقة المتناهية في دراسة ظواهرها بنفس قدر العلوم الطبيعية. فعالم الفيزياء أو الكيمياء مثلاً يبحث في ظواهر مادية يمكن التحكم فيها ودراستها بقدر كبير من الدقة بفضل الأدوات والأجهزة المقننة المتاحة لديه. وعلى الرغم من أنه لا يلاحظ بعض الظواهر الطبيعية ملاحظة مباشرة، مثل: الكهرباء أو المغناطيسية أو الصوت، إلا أنه يستدل على هذه الظواهر من أثرها في الأجسام، ومن ثم يقيس هذا الأثر باستخدام الأدوات والأجهزة الدقيقة المعدة لهذا الغرض.

وقد حذا علم النفس حذو العلوم الطبيعية، بل اهتم علماء النفس بالأساليب الكمية، ومنطق القياس اهتماماً أكبر، وذلك يرجع إلى أن الظواهر النفسية التي يودون قياسها أكثر تعقيداً وتشابكاً، ولا تزال أقل دقة في قياسها من الظواهر الطبيعية؛ فعالم البيولوجي يستطيع أن يثبت زهرة ويقوم بدراستها، وعالم الكيمياء يمكنه أن يتعامل مع المواد الكيميائية كما يشاء، وكذلك عالم الفيزياء. أما عالم النفس فمن الصعب عليه إخضاع الإنسان لتجاربه المختبرية، وتثبيت جميع المتغيرات التي يود عزل أثرها. كما أن الظواهر السلوكية يصعب قياسها على موازين مدرجة Scales كما هو الحال في موازين العلوم الطبيعية. فنحن بالطبع لا نستطيع قياس صفة المثابرة، أو الشجاعة، أو العدوانية أو العاطفة، بنفس القدر من الدقة، مثل: قياس أبعاد الحجرة، أو شدة الاستضاءة، أو كثافة سائل معين. وعلى الرغم من أن هناك ظواهر سلوكية أكثر بساطة ويسر في قياسها من ظاهرة المثابرة أو الشجاعة أو العاطفة، كظواهر إنسانية، إلا أن القياس النفسي والتربوي بعامة لا يعد يسيراً.

ومع هذا، فإن هناك تشابهاً كبيراً بين قياس الظواهر النفسية وقياس الظواهر الطبيعية. فنحن لا نرى المثابرة، أو الشجاعة، أو العاطفة لأنها خصائص مجردة، وإنما نستدل عليها من أثرها في سلوك الأفراد في مواقف معينة يمكن تحديدها تحديداً دقيقاً.

لذلك اهتم علماء النفس اهتماماً كبيراً بابتكار أساليب جديدة لقياس الظواهر النفسية، وطرقاً متنوعة لوصف واستقراء نتائج قياسها، والتعبير عنها بمقادير عددية، وذلك بهدف دراسة خصائص هذه الظواهر التي يمكن الاستناد إليها في تفسير السلوك والتنبؤ بحدوثه.

فهناك الآن العديد من الطرق والأساليب والأدوات العلمية التي يمكن عن طريقها قياس التحصيل والذكاء الإنساني، والاستعدادات، والميول، والاتجاهات، والشخصية، وغيرها. وبذلك يمكن تقويم الأفراد، واتخاذ قرارات معينة بشأنهم استناداً إلى نتائج هذا القياس العلمي في ضوء بعض المحكات.

وسوف أحاول في هذا الفصل، والفصول الثلاثة التالية التي يشتمل عليها هذا الباب إلقاء الضوء على أهم المفاهيم والمبادئ الأساسية للقياس النفسي والتربوي التي يستند إليها تصميم وبناء مختلف أدوات القياس التي سأتناولها في الباب الثاني.



مفهوم القياس وأهميته:

يعد قياس السلوك الإنساني مدخلاً أساسياً لفهم الفرد كإنسان متمايز وكعضو في مجتمع، وبذلك يجعل من الممكن التقييم الموضوعي لهذا السلوك، وعن طريق القياس نتوصل إلى معلومات يمكن الاستناد اليها في إصدار أحكام عن الوضع الراهن للأفراد والجماعات، وتقدير إمكانات الأفراد المستقبلية في مختلف مجالات السلوك الإنساني.

ومفهوم القياس Measurement له تعريفات متعددة، لعل أبسطها هو أنه وصف البيانات المتعلقة بخصائص الأشياء باستخدام الأعداد أو الجوانب الكمية في وصف سمات أو خصائص الأفراد (Guilford, 1954). ويرى سميث، وادمز (Smith & Adams, 1972) أن القياس بمعناه الواسع هو الجمع المنظم للمعلومات بترتيب معين، وهذا يتضمن عملية جمع وتنظيم المعلومات، وناتج هذه العملية.

ويعرف نانالي (Nunnally 1972) القياس بأنه يشتمل على قواعد تعيين أعداد للأشياء بحيث تمثل مقادير سمات هذه الأشياء. ويتفق هذا التعريف مع ما سبق أن قدمه ستيفنز (Stevens, 1951) من أن القياس هو تعيين أعداد أو رموز رقمية للأشياء أو الأحداث وفقاً لقواعد، فالقياس إذن يتطلب إجراءات أو عمليات معينة، استناداً إلى قواعد محددة تستخدم في المقارنة بين الأشياء أو الأحداث وفقاً لمعيار Standard أو ميزان Scale معرَّف تعريفاً دقيقاً.

وعلى الرغم من أن هذه القواعد تكون واضحة في بعض الحالات في القياس الفيزيائي مما لا يتطلب صياغات تفصيلية لها، كما في حالة قياس الطول، إلا أن هناك حالات كثيرة ربما تتطلب إجراءات معقدة، مثل قياس المقادير المختلفة لمكونات المركبات الكيميائية.

وقواعد قياس معظم السمات أو الخصائص النفسية ليست بسيطة أو مباشرة، مثل: ذكاء الأطفال ، ومقدار الاسترجاع في تعلم الترابط الثنائي، ومستوى الدافع، والاتجاه نحو جماعة معينة. غير أن القياس النفسي والتربوي يتبع نفس الخطوات المتبعة في القياس الفيزيائي. فنحن نُعرِّف السمة أو البعد المراد قياسه، ونحدد عمليات القياس والقواعد التي نستند اليها في القياس، وكذلك ميزان معين Scale لوحدات القياس يُعبِّر عن القياس الناتج.

ويمكن الإفادة من القياس الكمي في تحقيق ما يلي:

(أ) مرونة الوصف: فتطبيق إجراءات القياس تُمكِّننا من التمييز بين الأفراد، ووصف الفروق الفردية.

(ب) تيسير عملية التفسير: فالقياس لا يقتصر على تحديد ووصف الفروق الفردية، وإنما يُمكننا من ترتيب الأفراد بالنسبة للمتغير موضع القياس، وبذلك يمكن اشتقاق تفسيرات استناداً إلى الوضع أو الأداء النسبي لهؤلاء الأفراد في مقابل معيار معين.

(جـ) تحديد أنماط السلوك: فمرونة الوصف تُمكننا من تحديد وقياس الفروق الفردية في أنماط سلوكية مختلفة، حيث يمكن الربط بين هذه الأنماط بطرق مفيدة.

(د) اختزال البيانات: فتحديد ميزان القياس ووحداته يسمح باستخدام أساليب رياضية وإحصائية تلخص كميات كبيرة من البيانات، كما سيتضح في الفصل الثاني.

منطق السمات الإنسانية في القياس:

تبين مما سبق أننا لا نقيس الفرد وإنما نقيس خاصة أو سمة معينة لديه. فنحن نقيس طول المنضدة أو درجة حرارة الجسم، ونقيس ذكاء الطفل، والنضج الاجتماعي للمراهق. وعندما نتعامل مع ابسط الخصائص الفيزيائية مثل الطول، فإننا لا نتساءل عن معنى أو تعريف هذه الخاصة، حيث إن تعريف الطول محدد ومتفق عليه منذ زمن بعيد. ولكن هناك خواص فيزيائية كثيرة أخرى تتطلب الاتفاق على المقصود منها، مثل: المقاومة، والتحمل، والإجهاد، والمتانة، ...الخ. وعدم الاتفاق على معاني أو تعريفات هذه الخواص يجعل هناك اختلافاً في الإجراءات المناسبة لقياسها.

وتزداد حدة هذه المشكلة في علم النفس والتربية، حيث تفتقر كثير من السمات والخصائص الإنسانية إلى تعريفات محددة متفق عليها، وتختلف دقة هذه التعريفات باختلاف مدى تعقد السمة المعينة، ويؤثر ذلك بالطبع في مدى دقة قياس هذه السمات، فالقياس يتعلق بمظهر معين من مظاهر الأشياء المراد قياسها، حيث إننا لا نقيس الشيء وإنما نقيس بعض خصائصه وكذلك لا نقيس الفرد، وإنما نقيس سمة معينة لديه مثل: الذكاء، والقلق، والاتزان الانفعالي، والقيادة، ... الخ. ونعني بالسمة مجموعة مترابطة من السلوك الذي يحتمل حدوثه معاً. أي أن



السمة ليست صفة منفردة، وإنما تعد تكويناً فرضياً Construct، أو مفهوماً مجرداً، وليست شيئاً ملموساً. ولتوضيح ذلك نفترض أن طفلاً استطاع أن يحل مجموعة من التمارين الحسابية المتعلقة بالاستدلال العددي، والاستدلال اللفظي، والمتشابهات؛ فجميع هذه المظاهر السلوكية يمكن أن تندرج تحت "التفكير المجرد" أو "حل المشكلات" ولهذا فإننا ربما نستنتج أن هناك سمة عقلية تنظم أو تعكس هذه المجموعة المترابطة من السلوك الذي يمكن ملاحظته وتسجيله في مواقف اختبارية مقننة، ويمكن أن نطلق عليها سمة "الاستدلال المجرد "Abstract Reasoning". لذلك فإن السمات الإنسانية يمكن تعريفها في ضوء السلوك الذي يمكن ملاحظته ملاحظة مباشرة وتفسيره في إطارها. والاختبار الذي يُصمم لقياس سمة معينة ينبغي أن يتضمن مجموعة من المواقف التي تستدعي السلوك المناسب، ونستدل على موقع الفرد على متصل هذه السمة من استجابته لهذه المواقف.

وهذا يعني أن السمات الإنسانية لا تقاس قياساً مباشراً، وإنما يمكن الاستدلال عليها من أنماط سلوكية معينة، تنظمها هذه السمات. أي أن قياسها يكون غير مباشر، وهذا هو الحال دائماً في القياس النفسي والتربوي؛ فالسمات ينعكس أثرها في مجموعة من الأنماط السلوكية المترابطة التي يمكن ملاحظتها مباشرة في مواقف اختبارية مبتكرة. لذلك فإن تحديد هذه السمات، وبناء اختبارات لقياسها أو تقييمها ليست بالأمر الهين، فهو يعد عملاً فنياً وتقنياً متخصصاً، ويتطلب جهداً كبيراً.

طبيعة القياس النفسي والتربوي:

إن الأخذ بمفهوم السمة الذي ذكرناه يجعلنا نستنتج ما يلي:

(1) يمكن القول بأن القياس النفسي والتربوي نسبي وليس مطلقاً. وهذا يعني أن هناك متصل "Continuum" من السلوك نحاول عن طريق عملية القياس أن نحدد موضع الفرد بالنسبة لاقرانه على هذا المتصل. فقياس الذكاء أو الاستعدادات العقلية مثلاً يعتمد على تطبيق اختبار مناسب تم إعداده بطريقة معينة بحيث يشتمل على مهارات محددة يؤديها الفرد، مثل: حل مشكلات، استدلال، تعرف أشكال، عمليات حسابية ورياضية، معاني كلمات،...الخ، ومقارنة الدرجة التي يحصل عليها بمتوسط درجات أقرانه، بحيث نستطيع تحديد المركز النسبي للفرد في الذكاء أو الاستعدادات. لذلك

تجري بعض العمليات الإحصائية على الدرجات الخام لتحويلها إلى نوع من الدرجات ييسر هذه المقارنة، كما سيتضح في الفصل الثالث.

فالقياس النفسي والتربوي إذن يهدف لتكميم السمات والخصائص الإنسانية، أي تعيين أعداد أو قيم عددية تناظر سلوك أو أداء الفرد. وهذا يعني كما ذكرنا أن لدينا متصلاً نقيس عليه سمة معينة، وتهدف عملية القياس لتحديد موضع الفرد على هذا المتصل طبقاً لقواعد محددة. ومن الجدير بالذكر أن بعض أساليب وأدوات التقويم التربوي لا تؤدي إلى قيم عددية، أي أن نتائجها ليست كمية، ولذلك ربما لا يكون مناسباً اعتبار هذه النتائج قياسات.

(2) هناك مواقف يمكن أن يقارن فيها أداء الفرد بمحك أداء مطلق Criterion يتفق عليه أو يعد أساسياً، لكي يؤدي الفرد سلوكاً معيناً بطريقة سليمة، مثل: مهارة العزف على آلة موسيقية، أو مهارة الطبيب في إجراء عملية جراحية، أو مهارة اصلاح عطل معين في السيارة، أو غير ذلك. ويمكن أيضاً في مثل هذه الحالات مقارنة أداء الفرد بمتوسط أداء جميع أفراد المهنة المعنية. ولعل هذا المدخل يناسب بدرجة افضل القياس التربوي والمهني، بينما يناسب المدخل الأول القياس النفسي.

(3) على الرغم من تشابه عمليات القياس الفيزيائي والقياس النفسي والتربوي، إلا أن القياس الفيزيائي يتعامل بوحدات معيارية Measurement Units متفق عليها، مثل: الكيلومتر، والكيلوجرام، والفولت. أما القياس النفسي والتربوي فإنه يفتقر إلى مثل هذه الوحدات. فليس هناك وحدة للتحصيل في الحساب، أو وحدة للقدرة المنطقية الرياضية، أو وحدة للاستقرار الإنفعالي كما هو الحال في الوحدات الفيزيائية التي تتميز بوجود صفر مطلق، ووحدات متساوية على متصل ميزان القياس، كما سيتضح بعد قليل.

غير أن هناك مداخل متعددة للتوصل إلى وحدات قياس متسقة للسمات الإنسانية، مع تأكيد أن هذا الاتساق يعد أمراً نسبياً، حيث انه يوجد عدد قليل من وحدات القياس المتسقة المطلقة في مجالات العلوم النفسية والتربوية.

فالقياس النفسي بخاصة يستند إلى الأداء الملاحظ لمجموعة معيارية (نمطية) من



الأفراد، حيث يقارن أداء الفرد بأداء هذه المجموعة المعيارية، أي بمعيار الجماعة Group Norm، كما سيتضح في الفصل الثالث:

(4) على الرغم من أن تعريف ستيفنز Stevens للقياس الذي سبق أن أشرنا إليه يؤكد قواعد التكميم Quantification Rules ويعتبرها أساساً لعملية القياس، إلا أنه لم يحدد طبيعة هذه القواعد أو حدودها. وربما يرجع ذلك إلى أنه ينبغي التمييز بين كيفية القياس المستمدة من عملية القياس، وكيفية التحقق من صدق أدوات القياس، أو تحديد منفعتها. فهذه القواعد تعد افتراضات ينبغي التحقق من صحتها بالأساليب الإمبريقية في ضوء محكات خارجية عن عملية القياس ذاتها، حيث يتم تحديد مدى وجود علاقات وظيفية بين المقياس وبعض المتغيرات المهمة. وسوف أوضح ذلك بمزيد من التفصيل في الفصل الرابع.

ومن الجدير بالذكر أنه ينبغي تحديد مجموعة واضحة من القواعد استناداً إلى نموذج استدلالي معين، أو الخبرة السابقة، أو الحكم الصائب لمصمم الاختبار أو المقياس. ومحك اختبار هذه القواعد هو مدى فائدة المقياس في تفسير الظاهرة المستهدفة .

ومن هذا يتبين أن علم القياس النفسي والتربوي يواجه كثيراً من المشكلات والصعوبات كما هو الحال في بقية العلوم. غير أنه ربما يبدو أن هذه الصعوبات تزداد في العلوم النفسية والتربوية. وأحد الأسباب الرئيسة لذلك أن الإنسان الذي هو محور القياس في هذه العلوم دائم التغير، وأحياناً يصعب اخضاعه للضبط. فكما سبق أن أشرنا يستطيع عالم الكيمياء مثلاً التعامل مع معظم عينات دراسته من المواد الكيميائية بالطريقة التي يراها، وتظل هذه العينات في كثير من الأحيان ثابتة ومتسقة، ويمكنه تقسيمها أو خلطها. ولكن عينات الأطفال التي يقوم عالم النفس بدراستها يصعب بدرجة كبيرة أن تحقق هذا الاتساق وضبط الظروف. ومع هذا حدث تقدم ملحوظ في طرق تصميم المعالجات التجريبية للتغلب على ذلك، حيث أصبح من الممكن قياس مختلف السمات الإنسانية بأساليب متنوعة.

وينبغي تأكيد أن القياس مجرد أداة ووسيلة لغاية، وأهميته تكمن في درجة الإفادة من نتائجه في تحسين أداء الأفراد في مختلف مجالات الأعمال، وإثراء تعلم الطلاب في المؤسسات التعلمية. غير أنه تظل هناك حاجة إلى مزيد من الجهود المبتكرة لتطوير أساليب

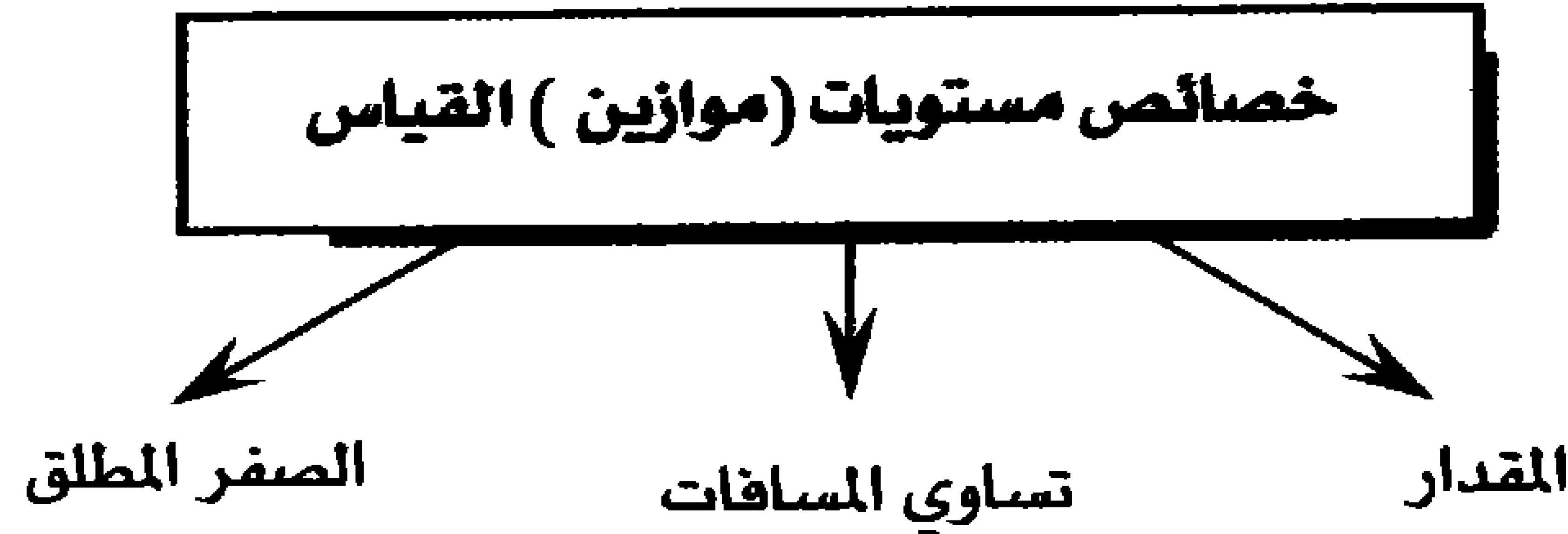
وأدوات القياس المتوافرة في وقتنا الحاضر، ومعاونة المشتغلين بالقياس، والمربين وغيرهم في استخدام الأدوات المتوافرة استخداماً أكثر فاعلية. ولعل البحوث السيكومترية والإديومترية المعاصرة، والتطورات التي أصبحت تميز علم القياس النفسي والتربوي، أسهمت اسهاماً ملحوظاً في هذا الشأن، ويتمثل ذلك في النظرية السيكومترية المعاصرة التي يُطلق عليها نظرية الاستجابة للمفردة الاختبارية Item Response Theory (IRT)، والقياس والتقويم البديل Alternative Assessment، حيث أفردت لكل منهما مرجعاً مستقلاً سبق نشره.

خصائص موازين القياس:

تختلف السمات والخصائص الإنسانية في مستويات قياسها؛ فخاصية نوع الفرد، أو مستوى تعليمه، أو حالته الاجتماعية، تختلف في مستوى قياسها عن خاصية ترتيب الطفل بين إخوته، أو ترتيبه في فصله، كما تختلف عن خاصية الذكاء، أو التحصيل، أو الميول. فالخصائص في الحالة الأولى تعد من المتغيرات الكيفية التصنيفية، وفي الحالة الثانية تعد من المتغيرات الرتبية، وفي الحالة الثالثة تعد من المتغيرات الكمية القياسية. وقد سبق أن ذكرنا أن القياس هو تعيين أعداد للسمات أو الخصائص طبقاً لقواعد معينة. ولكي تمثل هذه الأعداد السمة أو الخاصية المطلوبة يجب أن يكون هناك تناظراً احادياً بين بعض خصائص النظام العددي، والعلاقات بين مختلف مقادير الخاصية المراد قياسها. فإذا قمنا بتعيين أعداد تعبر عن السمات أو الخصائص المختلفة، فإننا نستطيع أن نتعامل مع هذه الأعداد بطرق رياضية معينة، ونتوصل من ذلك إلى استنتاجات يمكن أن نعيد تطبيقها على الظاهرة المقاسة. فالخصائص أو السمات الإنسانية يمكن وصفها عن طريق الأعداد بشرط أن يكون هناك تناظراً أحادياً بين هذه الخصائص أو السمات والنظام العددى المستخدم. فهناك خصائص معينة للأعداد ينبغي أن يوجد ما يناظرها في هذه الخصائص أو السمات المراد قياسها وهي: خاصية التفرد أو التميز، وخاصية الترتيب، وخاصية قابلية الجمع. وليس من الضروري أن يكون للسمة الإنسانية التي نُعين لها الأعداد الخصائص الثلاث السابقة لكي نتمكن من قياسها. إلا أن الاستفادة من استخدام الأعداد في القياس تعتمد على مدى توافر هذه الخصائص في السمة المراد قياسها..

وتعتمد مستويات أو موازين القياس Measurement Scales على ثلاث خصائص أساسية يوضحها الشكل التخطيطي (1-1) التالي:





شكل (1-1): يوضح مستويات (موازين) القياس

(1) المقدار: Magnitude: فالميزان Scale يحقق هذه الخاصية إذا كانت إحدى حالات سمة معينة يمكن الحكم بأنها أكبر أو أصغر أو تساوي حالة أخرى للسمة نفسها. فإذا استطعنا على ميزان الطول مثلاً أن نقول أن فرداً معيناً أطول من فرد آخر، فإن هذا الميزان يحقق خاصية "المقدار". أما إذا عينت أعداد لفرق كرة القدم للتعريف بكل فريق، فإن الاعداد في هذه الحالة لا تحقق هذه الخاصية. ولكن إذا كانت هذه الأعداد تُعبر عن ترتيب كل فريق بحسب عدد المباريات التي فاز فيها، فإن نظام الأعداد الجديد يحقق خاصية "المقدار".

(2) تساوي المسافات: Equal Interral: وهذه الخاصية تعني أن مقدار السمة الممثلة بوحدة قياس على الميزان تكون متساوية بغض النظر عن موقع هذه الوحدة على الميزان. وبعبارة أخرى تتحقق هذه الخاصية إذا كان الفرق بين نقطتين في أي موقع على الميزان تحمل نفس معنى الفرق بين نقطتين أخريتين تختلفان بعدد مساوٍ من وحدات الميزان. فالفرق بين 2 سم ، 6 سم على المسطرة لا يختلف معناه عن الفرق بين 8 سم، 12 سم، حيث إن هذا الفرق 4 سم في الحالتين، والسنتيمتر يعكس قدراً معيناً من الطول بغض النظر عن موقعه على المسطرة.

وعلى الرغم من أن خاصية تساوي المسافات تبدو بسيطة، إلا أنه يندر أن نجد دليلاً على أن الاختبار النفسي أو التربوي يحقق هذه الخاصية. فالفرق بين نسبتي الذكاء 60، 65 ليس له نفس معنى الفرق بين نسبتي الذكاء 115، 120، على الرغم من أن الفرق في كل من الحالتين 5 نقاط، وذلك لاختلاف مستوى الذكاء في كل منهما.

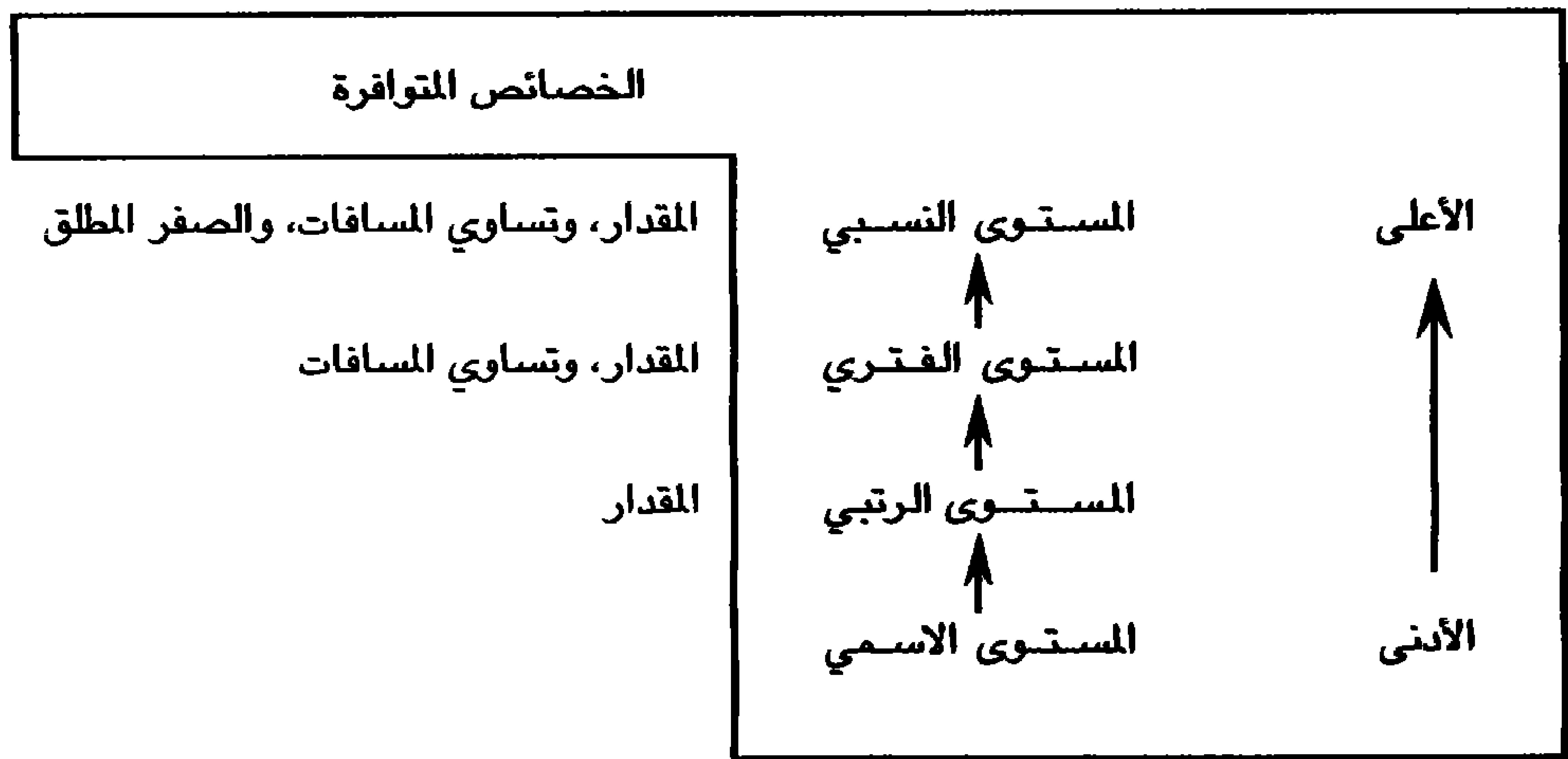
(3) الصفر المطلق Absolute Zero : تتحقق هذه الخاصية إذا انعدمت السمة موضع

القياس. فإذا كان عدد نبضات القلب صفراً، فإن هذا يعني أن القلب توقف عن العمل. أما اذا اعتبرنا أن أقل تقدير لسمة العدوانية (1) فإن هذا لا يعني انعدام السمة. وحتى إذا كان الصفر نقطة من نقاط ميزان قياس العدوانية، فإنه لا يعني بالضرورة انعدام العدوانية انعداماً كلياً، لذلك فإن هذا الميزان لا يحقق خاصية الصفر المطلق. والحقيقة أنه يصعب بدرجة كبيرة تعريف نقطة صفر مطلق لكثير من السمات النفسية.

مستويات أو موازين القياس:

يمكن التمييز بين أربعة مستويات أو موازين للقياس استناداً إلى الخصائص الثلاث السابقة، وتختلف هذه المستويات في تحقيقها لهذه الخصائص. لذلك فإن هذه المستويات هرمية بمعنى أن المستويات العليا تحقق خصائص جميع المستويات الأدنى منها بالإضافة إلى خصائص تميز المستوى المعين. وكل منها يمثل مستوى من مستويات التقدير الكمي للسمة المراد قياسها، كما يسمح بعمليات حسابية مختلفة.

ويوضح شكل (1 - 2) التالي هذه المستويات والخصائص التي تتحقق فيها:



شكل (1 - 2): مستويات القياس وخصائصها

وفيما يلي توضيحاً لكل من هذه المستويات الأربع:

المستوى الاسمي Nominal Scale: يعد أبسط مستويات القياس، حيث يقتصر على تصنيف الأفراد أو الأشياء في أقسام متنافية مثنى مثنى وفقاً للسمة المقاسة. ونعني بذلك أن



الفرد لا يجوز أن ينتمي إلى أكثر من قسم واحد في الوقت نفسه بالنسبة لسمة معينة، ولكن يمكن بالطبع تصنيفه في أكثر من سمة، كأن يكون ذكراً، ومستوى تعليمه جامعي، ومتزوج مثلاً.

أي أن القياس في هذه الحالة يتناول فروقاً كيفية أو نوعية، وعادة تشير الأعداد الناتجة إلى تكرارات للأفراد الذين ينتمون إلى كل من هذه الاقسام، وبالتالي لا نستطيع إجراء أي من العمليات الحسابية الأربع على هذه الأعداد لأنها تُعبر عن تكرارات لأقسام مختلفة النوع، أي منفصلة، وعملية تحديد التساوي هي العملية الأساسيـــة في هذا المستوى، حيث يمكن فقــط تحديد أن جميع الأفـراد الذين ينتمون إلى قسم معين يمتلكون خاصة مشتركة لا يمتلكها من لا ينتمون إلى هذا القسم. لذلك فإن الميزان الاسمي في القياس محدود الفائدة في علم النفس والتربية، نظراً لأنه يسمح فقط بتصنيفات عريضة دون تحديد أي مقدار كمي وهو ما يهمنا في القياس، حيث إنه يعتمد على عملية العد Counting، أي تحديد عدد الأفراد الذين ينتمون إلى قسم معين. وبذلك تتضح الفروق الكيفية بين الأفراد، وليست الفروق الكمية.

المستوى الرتبي Ordinal Scale: يعتمد هذا المستوى على ترتيب الأفراد في سمة أو خاصة معينة من الأكبر إلى الأصغر، وتدل الأعداد الناتجة على الوضع النسبي للأفراد في هذه السمة. فهذا الطالب ترتيبه الأول في اللغة العربية بالنسبة لطلاب فصله، وهذا ترتيبه الثاني، وهكذا. وبالطبع لا تتساوى بالضروري الفروق بين درجات الطلاب من ذوي الرتب المتتالية. وكثير من السمات أو الخصائص الإنسانية الكيفية يمكن ترتيبها بهذا الشكل، لذلك فإن هذه الأعداد الرتبية لا تدل على مقدار مطلق للسمة لدى الفرد في المجموعة. ونظراً لصعوبة تحقق تساوي المسافات وتوافر صفر مطلق في كثير من الظواهر النفسية والتربوية، فإن المستوى الرتبي شائع الاستخدام في قياس هذه الظواهر، لأنه يحقق فقط خاصية "المقدار".

المستوى الفتري Interval Scale: يتميز هذا المستوى بأنه يحقق خاصيتي المقدار، وتساوي المسافات على ميزان قياس سمة أو خاصة معينة. فالترمومتر مقسَّم إلى وحدات متساوية، والفرق بين درجتي الحرارة 35، 37 يساوي الفرق بين الدرجتين 39 ، 41 ، أو بين أي درجتين متجاورتين أخريتين.

ومما هو جدير بالذكر أن معظم الاختبارات المستخدمة في قياس الذكاء، أو التحصيل، أو

الاستعدادات، أو الميول، أو الشخصية، تعتمد على هذا المستوى. إذ يمكن اعتبار الفرق بين نسبتي الذكاء 80، 100 في اختبار للذكاء يكافىء الفرق بين نسبتي الذكاء 100 ، 120 أو 120، 140 بشروط معينة، وهكذا.

غير أنه لا يجوز أن نقول أن نسبة الذكاء 80 نصف نسبة الذكاء 160، أو أن تحصيل الطالب الذي حصل في اختبار تحصيلي مقنن في الحساب على الدرجة 90 ضعف تحصيل طالب آخر حصل في الاختبار نفسه على الدرجة 45، وذلك لعدم وجود "صفر مطلق" على تدريج المقياس، إذ لا يوجد فرد ذكاؤه صفراً، أو تحصيله صفراً. لذلك فإن نقطة الصفر على هذا التدريج تكون اعتبارية كما هو الحال في نقطة الصفر على التدريج العددي للترمومتر، إذ أن هذا الصفر لا يعني انعدام الحرارة وإنما هو مستوى حرارة اتفق علماء الفيزياء على استخدامه كنقطة بدء قياس هذه الظاهرة على هذا التدريج. ويمكن أن يناظر الصفر الاعتباري في الاختبارات والمقاييس النفسية والتربوية متوسط توزيع درجاتها. ويهتم علماء القياس بالحصول على ميزان فتري في الاختبارات النفسية والتربوية لأنه يُمكننا من تحويل الدرجات تحويلاً خطياً Linear Transformation في هذه الحالة، أي تتحول الدرجات من ميزان إلى ميزان آخر وحداته مختلفة، كما هو الحال عند تحويل درجات الحرارة المئوية تحويلاً خطياً إلى درجات فهرنهيتية أو العكس. وهذا يعني أن درجات الاختبارات أو المقاييس المختلفة يمكن وضعها على ميزان مشترك Common Scale، وبالتالي يمكن مباشرة مقارنة أداء الأفراد على مقائيس مختلفة. كما يمكن استخدام معظم الأساليب الاحصائية التي تفترض أن يكون ميزان القياس فترياً كما سيتضح في الفصل الثاني.

المستوى النسبي Ratio Scale: يعد أدق مستويات القياس، حيث إن هذا المستوى يحقق الخصائص الثلاث السابق ذكرها، وهي: المقدار، وتساوي المسافات، والصفر المطلق الذي يمثل انعدام السمة المقاسة. ولعل المسطرة العادية تعد مثالاً بسيطاً للميزان النسبي، حيث إن الفروق بين التداريج المدونة عليها (السنتميترات أو البوصات) متساوية في الطول، وتبدأ هذه التداريج بصفر مطلق يمثل انعدام الطول، وعندئذ نستطيع القول بأن الطول 15سم مثلاً نصف الطول 30 سم. لذلك فإن هذا المستوى يسمح بإجراء العمليات الحسابية الأربع على الأعداد الناتجة عن القياس.

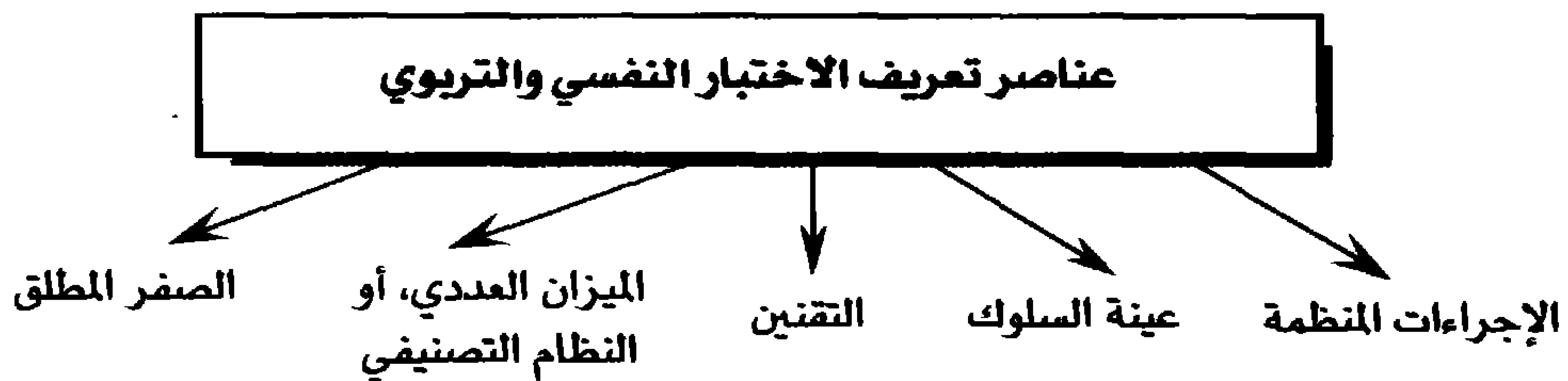
وعلى الرغم من أن معظم الخصائص الفيزيائية للأفراد مستوى قياسها نسبي، إلا أنه يندر استخدام هذا المستوى في القياس النفسي والتربوي لعدم توافر صفر مطلق يناظر حقيقة انعدام الظاهرة أو السمة المقاسة. لذلك فإن مستوى قياس السمات والخصائص النفسية والتربوية يكون عادة فترياً أو رتبياً.



ولعل عرضنا لمستويات القياس الأربعة التي توضح الطرق المختلفة لتطبيق الموازين في تعرّف سمات الأفراد والفروق بينهم يجعل أخصائي القياس، ومستخدم الاختبارات والمقاييس مدركاً أن المستوى النسبي الذي يتميز بتساوي الفترات، وتوفر الصفر المطلق ليس هو المستوى الوحيد للقياس، حيث إن كثيراً من السمات النفسية والتربوية لا تحقق متطلبات هذا المستوى. كما أن تصنيف هذه المستويات في مجموعات أربع يسمح بوضع الاختبارات والمقاييس في ترتيب هرمي من حيث درجة تعقد القياس ودقته. وبذلك يمكن أن يتخذ قراراً فيما يتعلق بالأدوات المتنوعة، واختيار الأداة الأكثر ملاءمة للغرض من القياس.

مفهوم الاختبار كأداة قياس:

تعد الاختبارات النفسية والتربوية أدوات قياس للسمات الإنسانية شأنها شأن أدوات قياس الظواهر الطبيعية، مثل الترمومتر. غير أن كثيراً من الناس يعتقدون أن الاختبارات مرادفة للامتحانات المدرسية. وهذه تعد نظرة ضيقة ومحدودة لمفهوم الاختبار Test الذي يُعد مفهوماً متسعاً، ويتضمن معظم المفاهيم والمبادىء التي سبق توضيحها. وتشير أدبيات القياس إلى تعريفات متعددة لهذا المفهوم. إذ ترى انستازي (Anastasi, 1976) أن الاختبار هو أداة قياس موضوعية ومقننة لعينة من السلوك. ويرى براون (Brown, 1970) أن الاختبار هو إجراء منظم لقياس عينة من سلوك الفرد. بينما يرى شيس (Chase, 1978) أن الاختبار هو أسلوب منظم لمقارنة أداء الفرد بمعيار أو مستوى أداء محدد. أما كرونباك (Cronbach, 1970) فيرى أن الاختبار النفسي أو التربوي هو أسلوب منظم لملاحظة ووصف خاصية أو اكثر من خصائص الفرد استناداً إلى ميزان عددي أو نظام تصنيفي. وعلى الرغم من الاختلاف الظاهري لهذه التعريفات إلا أنه توجد بينها عناصر مشتركة يوضحها الشكل التخطيطي (1 - 3) التالي:



شكل (1 - 3) يوضح عناصر تعريف الاختبار النفسي والتربوي

وفيما يلي توضيحاً موجزاً لكل من هذه العناصر:

الإجراءات المنظمة Systematic Procedures : فالاختبار ينبغي أن يستند في بنائه وتطبيقه وتصحيحه أو تقدير درجاته إلى قواعد واضحة ومحددة. فمفردات الاختبار أو المهام التي يشتمل عليها لا يتم انتقاء محتواها من النطاق السلوكي المراد قياسه بأسلوب منظم فحسب، بل تُطبق أيضاً جميع مفرداته أو مهامه على جميع الأفراد المختبرين في ظروف متكافئة.

عينة السلوك Behavioral Sample : فالمفردات أو المهام التي يشتمل عليها الاختبار ينبغي أن تعكس السلوك الذي يمكن الاستدلال منه على السمة المراد قياسها، وهذا يتطلب أن تكون عينة مفرداته ممثلة لجميع مكونات هذه السمة.

التقنين Standardization : فالاختبار المقنن هو ذلك الاختبار الذي تتوحد فيه وتتحدد بدقة طريقة تطبيقه، ومواده، وتعليمات وزمن إجابته، وطريقة تصحيحه أو تسجيل درجاته، بحيث يصبح الموقف الاختباري موحداً بقدر الإمكان لجميع الأفراد. وهذا يجعل من الممكن المقارنة بين درجات الأفراد الذين طُبق عليهم الاختبار.

الميزان العددي، أو النظام التصنيفي Scale or Classification system : فبعض الاختبارات مثل اختبار الذكاء وغيره من اختبارات الجوانب المعرفية يؤدي إلى قياسات تحقق المتطلبات التي تضمنها تعريف القياس الذي سبق أن اوضحناه، بحيث يمكن وصف السلوك المراد قياسه كمياً. غير أن هناك أساليب أو أدوات قياس أخرى، مثل المقابلة Interview يستخدمها الأخصائيون النفسيون في تصنيف سلوك الفرد وفقاً لنمط استجاباته. فاختبارات الجوانب المعرفية تستخدم موازين عددية تحدد مكانة الفرد بين أقرانه في السمة المقاسة، وذلك بتحديد نسبة عدد الأفراد الذين يتفوق عليهم الفرد في هذه السمة، في حين أن أسلوب المقابلة غير المقننة مثلاً يستخدم نظاماً تصنيفياً استناداً إلى مفاهيم نفسية تحليلية تصف استجابات كل فرد أجريت له المقابلة. كما أن هناك اختبارات تربوية، مثل: بعض الاختبارات التحصيلية، واختبارات الأداء التي تقيس مهارات إجرائية محددة لا يقارن فيها أداء الفرد بأداء أقرانه، وإنما بمستوى أداء، أو محك كفاءة مناسب لمعرفة مدى تمكن الفرد من هذه المهارات، وتعرُّف المهارات التي لم يستطع تحقيقها. وهذا يعني أن هناك أطراً مرجعية مختلفة لتفسير درجات الاختبار، والغرض من الاختبار أو المقياس هو الذي يحدد الأساس الذي في ضوئه يُختار الإطار المرجعي المناسب.

الموضوعية: Objectivity: فتطبيق الاختبار، وتصحيحه، وتفسير درجاته يكون موضوعياً إذا لم يتأثر ذلك بالأحكام الذاتية للمختبر. ولعل هذا يرتبط بعملية تقنين الاختبار التي سبق توضيحها. فالفرد الذي يُطبق عليه الاختبار أو المقياس ينبغي ألا تختلف الدرجة التي يحصل عليها باختلاف الشخص القائم بعملية التصحيح. وهذه الموضوعية التامة تعد هدفاً يسعى إليه القائمون بإعداد الاختبارات والمقاييس التربوية والنفسية على الرغم من صعوبة تحقيق ذلك في الممارسات الفعلية. ومع هذا فإن كثيراً من الاختبارات والمقاييس المتوافرة تحقق ذلك بدرجة جيدة (Anastasi, 1970). وخلاصة ذلك أن الاختبار الذي تُستخدم في بنائه إجراءات موضوعية منظمة، ويحدد الميزان العددي أو النظام التصنيفي الذي يستند اليه تفسير درجاته، يؤدي إلى تقليل تأثير المتغيرات الموقفية والشخصية إلى أقل حد ممكن، ويسمح بمقارنة الفرد بأقرانه أو بمحك أداء محدد.

أنواع الاختبارات والمقاييس النفسية والتربوية:

تختلف أدوات القياس النفسي والتربوي وفقاً لكثير من الجوانب أو الأبعاد مثل: المحتوى، والغرض، وطريقة التصميم والبناء، وطريقة التطبيق، ونمط الأداء، وكيفية الاستجابة، وطريقة التصحيح، وكيفية تفسير الدرجات. وفيما يلي توضيحاً لكل من هذه الجوانب:

المحتوى: تختلف الاختبارات والمقاييس في محتواها أو المهام التي تشتمل عليها؛ فبعضها يكون محتواه لغوياً، مثل: المفردات والجمل، والبعض الآخر يكون محتواه غير لغوي، مثل: المسائل الحسابية والجبرية، والأشكال والخرائط، والصور، والرسوم، وغير ذلك. وتختلف أيضاً الاختبارات التحصيلية في محتواها، فبعضها يكون محتواه محدوداً حيث يقتصر مثلاً على مصطلحات جغرافية، أو صيغ رياضية، أو رموز كيميائية، والبعض الآخر يكون محتواه متسعاً حيث يشتمل على تعبيرات لغوية، ومسائل، ونصوص، وغير ذلك. كما توجد اختبارات تشتمل على مواد عيانية، مثل: المكعبات، والمتاهات، والأجهزة وغيرها، وتتطلب أن يتعامل الفرد معها يدوياً، وتسمى اختبارات الأداء Performance Tests، وبعض هذه الاختبارات تناسب الأطفال الصغار، والأميين.

الغرض: يمكن التمييز بين أدوات القياس من حيث طريقة الفرد في الأداء، ففي الاختبارات التي تقيس الجوانب المعرفية لدى الفرد كالتحصيل، والذكاء، والاستعدادات تُعطي تعليمات

للمختبرين بأن يبذلوا أقصى جهدهم للحصول على أعلى درجة ممكنة. لذلك يسمى هذا الأداء بالأداء الأقصىMaximum Performance. أما الاختبارات والمقاييس التي تقيس الجوانب غير المعرفية، أي الوجدانية، مثل: الميول والاتجاهات والشخصية، فإنها تهتم بقياس أداء الفرد الذي يقوم به في نشاطه اليومي المعتاد، أي السلوك النمطي الذي يميز هذا الفرد، مثل: هل يحب تنسيق الزهور؟ أو هل يحبذ الاختلاط في الجامعة؟ أو هل هو قلق باستمرار؟ والأداء الذي تتطلبه هذه الاختبارات والمقاييس يسمى الأداء المميز Typical performance، وتشتمل هذه الاختبارات عادة على عبارات أو فقرات تتطلب أن يجيب عليها الفرد بنعم أو لا، أو موافق أو غير موافق، ولا يوجد لأي عبارة منها إجابة صحيحة واحدة، بل تعتمد على الرأي الشخصي للفرد الذي يعكس سلوكه المميز في مواقف معينة.

طريقة التصميم والبناء: بعض الاختبارات والمقاييس تكون مقننة Standardized Tests، والبعض الآخر يكون غير مقنن. وقد سبق أن أوضحنا مفهوم التقنين. فالاختبار أو المقياس المقنن هو ذلك الاختبار الذي تكون طريقة تطبيقه ومواده وتعليمات إجابته وطريقة تصحيحة موحدة قدر الامكان لجميع المختبرين. ويمر بناء الاختبارات المقننة بعامة بمراحل متعددة، وتتطلب خبرة عدد من الاختصاصيين في القياس والتقويم التربوي والنفسي، وتجريب الاختبار على عينة عشوائية ممثلة لمجتمع الأفراد الذي أعد من أجله، والتوصل إلى ما يسمى معايير الاختبار Test Norms التي تفسر درجة الفرد بالنسبة لأقرانه استناداً اليها. وسوف أوضح هذه المعايير في الفصل الرابع.

ويلاحظ أن معظم اختبارات الذكاء والاستعدادات تكون مقننة، وكذلك توجد اختبارات تحصيلية مقننة تتعلق بمجالات دراسية متعددة. وتتميز هذه الاختبارات بأن محتواها متسع، بمعني أنها تشمل معظم الأهداف العريضة المتعلقة بهذه المجالات، وتقيس عادة معلومات أتفق بوجه عام على أنها ضرورية لمتوسطي الأداء في فرقة دراسية معينة. أما الاختبارات غير المقننة فتشمل بعض أدوات القياس النفسي، مثل: المقابلة الشخصية، والمقابلة الكلينيكية، والمقابلة لدراسة تاريخ الحالة، وغيرها. كما تشمل الاختبارات التحصيلية الصفية التي يكتبها المعلم لتلاميذه وترتبط ارتباطاً مباشراً بما ينتهي المعلم من تدريسه لهم، ولا تتطلب إجراء دراسات استطلاعية على عينات ممثلة من طلاب المدارس المختلفة في الفرقة المستهدفة، كما لا تتطلب تكوين جداول معايير، وإنما تتطلب عناية بتحديد الأهداف الإجرائية للوحدات الدراسية



التي يهدف المعلم لتحقيقها لدى تلاميذه، ووضع مواصفات دقيقة للاختبار بحيث تقيس مفرداته الأهداف المحددة تحديداً مباشراً.

طريقة التطبيق: تُطبق بعض الاختبارات والمقاييس على فرد واحد في وقت واحد، مثل: اختبارات الذكاء الفردية Individualized Tests كاختبار بينيه Binet، واختبار ويكسلر Wechsler، ويقوم بتطبيقها عادة شخص مدرب على تطبيق الاختبار، وتصحيحه وتفسير نتائجه.

وبعض المقاييس النفسية الفردية، مثل: اختبار بقع الحبر لرورشاخ Rorshach الذي يعد من الاختبارات الإسقاطية المستخدمة بكثرة في قياس الشخصية الإنسانية، واختبار ويكسلر للذكاء يحتاج تطبيقها إلى أخصائي نفسي، حيث تسمح له بأقصى قدر من التفاعل بينه وبين الفرد المختبر، وتُقدم فرصاً ثرية للملاحظات الكلينيكية التي تفيد في تشخيص الاضطرابات النفسية، والتخلف العقلي أو الدراسي، وغير ذلك. فالأخصائي يمكنه ملاحظة المدخل الذي يستخدمه المختبر وأدائه في المهام الاختبارية، ويتدخل الأخصائي أحياناً بتساؤلات تفيد في متابعة استجابات الفرد وتوضيحها وفهمها بدرجة أفضل. غير أن هذا النوع من الاختبارات يتطلب ظروفاً معينة، مثل توافر حجرة خاصة يُجري فيها الاختبار، كما يتطلب وقتاً وكلفة لتقييم كل فرد مختبر.

أما النوع الآخر فهو الاختبارات الجماعية Group Tests التي تعتمد عادة على الورقة والقلم، وتُطبق على مجموعة كبيرة من الأفراد في وقت واحد. وهي شائعة الاستخدام في الصفوف المدرسية العادية لأغراض التوجيه التعليمي أو المهني، وكذلك في المؤسسات العسكرية لانتقاء الجنود للمهام المختلفة. وتتميز هذه الاختبارات بأنها أكثر فعالية من الاختبارات الفردية في تطبيقها، وأقل كلفة، كما أن تطبيقها لا يتطلب شخصاً مدرباً. غير أن تفسير الدرجات المستمدة من كل من نوعي الاختبارات يتطلب توافر معارف ومهارات إحصائية معينة لدى القائمين بذلك.

نمط الأداء: نقصد بنمط الأداء نوع النشاط الذي يؤديه الفرد في الاختبار أو المقياس. فبعض الاختبارات تتطلب الكتابة، أي الورقة والقلم، مثل: الاختبارات التحصيلية، واختبارات الذكاء والاستعدادات، وبخاصة الجماعية منها، ومقاييس الجوانب الوجدانية.

وتنقسم اختبارات الورقة والقلم إلى اختبارات تعتمد على التعرف Recognition، وأخرى تعتمد على الاسترجاع Recall .

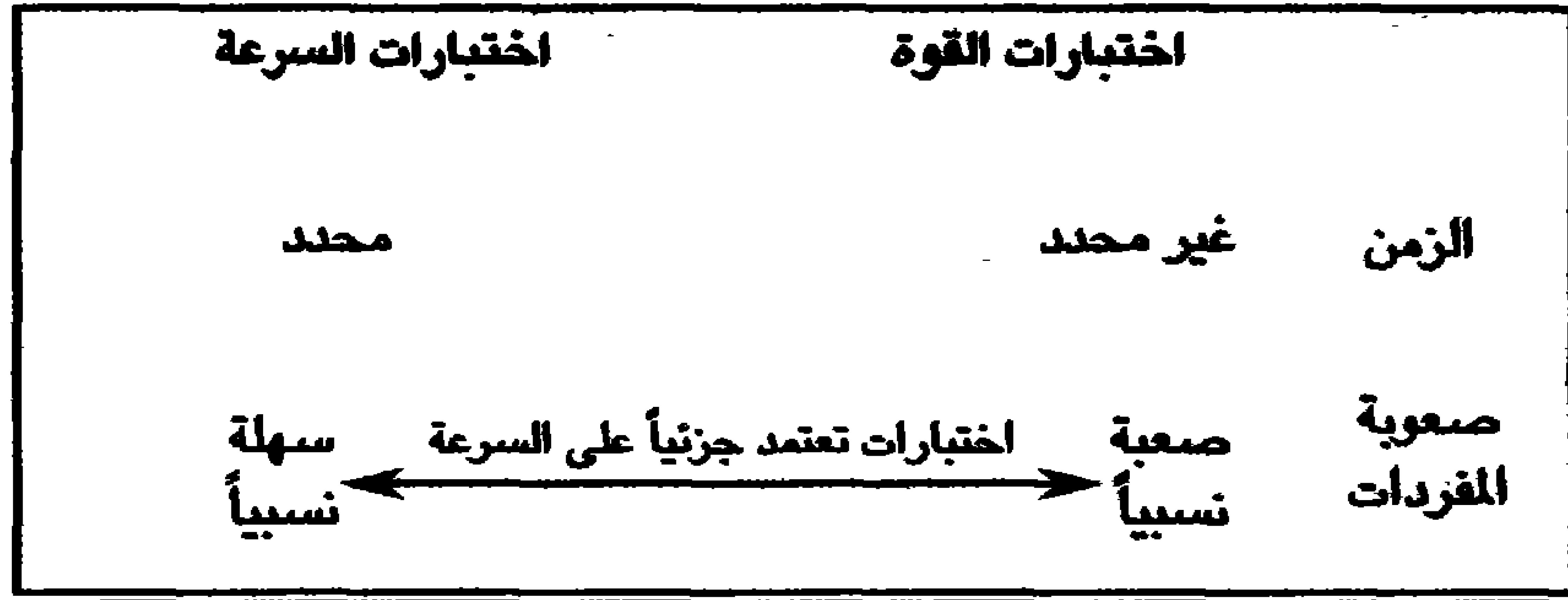
ففي اختبارات التعرف يقدم للفرد سؤالاً أو عبارة بالاضافة إلى مجموعة من الإجابات أو الخيارات التي يختار منها الإجابة الصحيحة، أو أفضل الإجابات، أو ما ينطبق عليه بدرجة أفضل. أما في اختبارات الاسترجاع فيُقدم السؤال أو العبارة دون خيارات معينة، وعلى الفرد أن يستجيب بأي اجابة يراها مناسبة في ضوء المحكات المعطاة.

وبعض آخر من الاختبارات يتطلب الأداء العملي، مثل: بعض اختبارات التربية الفنية والموسيقية، وبعض اختبارات الذكاء الفردية وبخاصة التي تطبق على الأطفال، والمقاييس الوجدانية.

كيفية الاستجابة: يعتمد بعض الاختبارات النفسية والتربوية على السرعة في الاستجابة Speed Tests، والبعض الآخر يعتمد على القوة Power Tests. فاختبار السرعة يكون له زمن محدد تحديداً دقيقاً ينبغي أن يلتزم به الفرد أثناء إجابته على الاختبار. وتتميز مفردات هذا النوع من الاختبارات بالسهولة، بحيث إننا نتوقع عدداً قليلاً من الأفراد يخطئون في بعض المفردات. ولكن لا يستطيع الانتهاء من إجابة جميع مفردات الاختبار الا بعض الأفراد مرتفعي القدرة. ومثال ذلك: بعض اختبارات الذكاء والاستعدادات الكتابية، والاختبارات التحصيلية التي تقيس معارف ومهارات أساسية. وتعد درجة الفرد في هذا النوع من الاختبارات مؤشراً لسرعة إجابته.

أما اختبارات القوة فإنه لا يكون لها زمن محدد للإجابة، ويُعطي الفرد فرصة كافية للإجابة قدر الإمكان على جميع المفردات. وتميل مفردات هذا النوع من الاختبارات إلى الصعوبة، وتتباين في مستوى صعوبتها. ومثال ذلك: بعض اختبارات الذكاء والاستعدادات، واختبارات الكفايات. ويمكن اعتبار الاختبارات التي تعتمد بدرجة أساسية على السرعة أو القوة ممثلة بطرفي متصل لتباين زمن الاجابة، كما هو موضح بالشكل (1 - 4) التالي:





شكل (1 - 4): يوضح متصل اختبارات السرعة والقوة

ويتضح من شكل (1 - 4) أنه على الرغم من اعتماد بعض الاختبارات اعتماداً اساسياً على السرعة أو القوة، إلا أن هناك اختبارات تعتمد على السرعة اعتماداً جزئياً Partially Speeded Tests، أي تتطلب السرعة والدقة معاً.

طريقة التصحيح: تختلف الاختبارات والمقاييس التربوية والنفسية في طريقة تصحيحها، فبعضها يعتمد على طريقة موضوعية في التصحيح Objective Test، وبعض آخر يعتمد على طريقة ذاتية Subjective Test . ونقصد بموضوعية التصحيح عدم تأثر درجات الفرد بالقائم بعملية التصحيح بغض النظر عن نوع المفردة أو محتواها. فإذا كان محك التصحيح واضحاً ومتفقاً عليه، فإنه يمكن التقليل من الأحكام الذاتية قدر الامكان. ولعل مفردات الاختيار من متعدد، والصواب أو الخطأ، والمزاوجة، تعد من المفردات الموضوعية حيث إنه يجري تصحيحها عادة باستخدام مفتاح تصحيح Scoring Key. أما أسئلة المقال فإن تصحيحها ربما يختلف باختلاف القائمين بتقدير الدرجات، مما يجعل التصحيح متأثراً بالذاتية. غير انه يمكن العنأية بإعداد محكات تفصيلية محددة يُسترشد بها في عملية التصحيح، مما يسهم في زيادة موضوعية تقدير درجات اسئلة المقال.

كيفية تفسير الدرجات: يمكن تفسير درجات الاختبارات والمقاييس إما عن طريق موازنة أداء كل منهم بأداء اقرانه في السمة التي يقيسها الاختبار بهدف ترتيبهم بالنسبة لبعضهم البعض الآخر في هذه السمة، أو عن طريق موازنة أداء كل منهم بمحك أو مستوى أداء محدد بغض النظر عن المركز النسبي لكل منهم في السمة موضع القياس.

وعلى الرغم من انه يمكننا تصميم وبناء اختبار أو مقياس تُفسَّر الدرجات المستمدة منه بالطريقتين معاً، إلا أنه يحسن التمييز بين كل من التفسيرين. فإذا كان الهدف من بناء

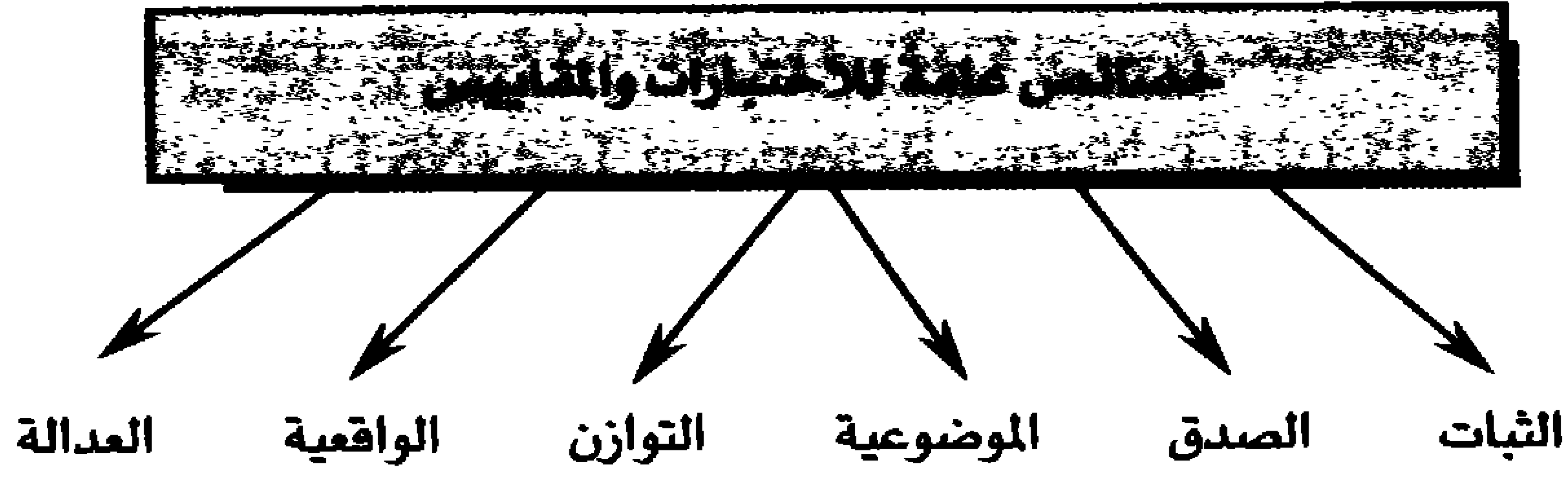
الاختبار تقدير أداء الفرد بالنسبة لأداء أقرانه في قدرة أو سمة معينة، فان الاختبار يسمي اختبار مرجعي المعيار Norm - Referenced Test (NRT). أما إذا كان الهدف هو تقدير أداء الفرد بالنسبة لمحك أو مستوى أداء معين (أي نسبة معينة من الأهداف أو المستويات التعليمية Standards المحددة لمقرر دراسي معين) فإن الاختبار يسمى اختبار مرجعي المحك Criterion - Refernced Test (CRT).

ويمدنا النوع الأول من الاختبارات بتفسير لدرجة الفرد في ضوء معيار الجماعة Group Norm، أي في ضوء متوسط أداء الجماعة، أو في ضوء نسبة مئوية من الأقران الذين ينتمي إليهم فرد معين يتفوق عليهم. وتصبح الدرجة لا معني لها ولا دلالة ويصعب تفسيرها إذا أغفلنا ردها إلى هذا المعيار. وتعد الاختبارات التحصيلية المقننة، واختبارات الذكاء والاستعدادات، وبعض مقأييس الجوانب الوجدانية من هذا النوع.

أما النوع الثاني فيمدنا بمعلومات تفصيلية محددة عن تحصيل الطالب، وكفأياته، وما انجزه من الأهداف أو المستويات المتعلقة بمقرر دراسي معين. لذلك فإن الخطوة الأساسية في بناء هذا النوع من الاختبارات هي تحديد الأهداف والمستويات أو التوقعات السلوكية التي يقيسها الاختبار، وتصبح هذه الأهداف أو التوقعات بمثابة المحك الذي يُنسب إليه أداء الفرد، وتفسر درجته في ضوئها لمعرفة المستوى التحصيلي الذي حققه، والأهداف التي لم يستطع تحقيقها. لذلك فإن هذه الاختبارات تتميز بأن لها قيمة تشخيصية في تحديد جوانب القوة والضعف في تحصيل كل طالب بالنسبة لمجموعة محددة من المعارف والمهارات. كما أنها تساعد في تحديد المستوى المدخلي للطالب Entry Behavior الذي يُمكنه من البدء في دراسة وحدة تعليمية معينة، أو تعلم مهارة مرجوة، وكذلك في تحديد الصف المناسب للطالب المنقول، أو الذي انقطع عن الدراسة لظروف معينة.

خصائص عامة للاختبارات والمقاييس النفسية والتربوية:

إن الاختبارات والمقاييس التربوية والنفسية هي أدوات قياس عينة مقننة من السلوك بطريقة منظمة كما أوضحنا فيما سبق. ولكي يفي الاختبار أو المقياس بهذه الشروط ينبغي أن تتوافر فيه مجموعة من الخصائص يوضحها الشكل التخطيطي (1- 5) التالي:



شكل (1 - 5) يوضح خصائص عامة للاختبارات والمقاييس

وفيما يلي توضيحاً لكل من هذه الخصائص:

(1) **الثبات Reliability** : ويعني اتساق الدرجات التي يحصل عليها الفرد في الاختبار أو المقياس. واتساق القياس يمكن الحكم عليه من حيث: مفردات الاختبار، وزمن تطبيق الاختبار، والقائمين بالعملية الاختبارية، والمختبرين. وربما تتضح خاصية الثبات عندما نستخدم المسطرة في قياس طول خط مستقيم، فلو وجدنا أن طوله 6.5 سم، ثم أعدنا قياسه مرة اخرى بالمسطرة نفسها ووجدنا أن الطول أصبح 6.7 سم، فإننا بلا شك سوف نفكر في سبب وجود هذا الفرق في القياس. فربما نقول أن المسطرة بها عيب معين، أو أن اليد قد اهتزت أثناء عملية القياس، أو غير ذلك من الأسباب؛ فعندئذ نستنتج أن العدد المناظر للطول الذي حصلنا عليه غير ثابت، أو غير متسق. وهذا يقلل من الثقة في القياس، ويجعل الاعتماد عليه مضللاً. كذلك الحال في القياس النفسي والتربوي، إذ لا بد أن تكون البيانات المستمدة من أدوات القياس ثابتة حتى يمكن الاعتماد عليها في اتخاذ قرارات صائبة تتعلق بالأفراد. وهذا يتطلب التحقق من خاصية الثبات قبل استخدام أداة قياس معينة.

(2) **الصدق Validity**: فأدوات القياس تُستخدم في قياس سمة معينة لدى الفرد وفقاً لتعريف محدد لهذه السمة، فإذا كانت هذه الأداة تقيس هذه السمة بحق، فإن البيانات المستمدة منها تفيد في اتخاذ قرارات مناسبة تتعلق بغرض معين. وتشتمل أدوات القياس على عينة من الأسئلة أو المفردات أو العبارات المتعلقة بمكونات سمة معينة، ونحاول من هذه العينة الاستدلال على أداء الفرد في مواقف معينة خارج حدود الموقف الاختباري. فنحن لا نقيس الذكاء باختبار ذكاء معين لمجرد القياس، وإنما نود

الاستدلال على سلوك الفرد وأدائه في مواقف تعليمية أو مهنية خارجة عن نطاق المواقف التي طبِّق فيها الاختبار. اذ ربما نود التنبؤ بنجاحه في الدراسة في إحدى الكليات مثلاً استناداً إلى درجة ذكائه التي حصل عليها في الاختبار.

فصدق الاختبار إذن يعني مدى استفادتنا من الدرجات التي يحصل عليها الأفراد في اختبار أو مقياس نفس أو تربوي معين في اتخاذ قرارات مناسبة تتعلق بموقف أو مواقف معينة بقدر كبير من الثقة.

(3) الموضوعية **Objectivity**: فدرجات الاختبارات ينبغي ألا تتأثر قدر الإمكان بعوامل تختلف عن السلوك المحدد المراد قياسه، مثل: غموض الأسئلة أو الاستجابات، وأخطاء التصحيح البسيطة، ومزاج القائم بالتصحيح. وكذلك ربما تتأثر الدرجات بوجود انطباعات مسبقة فيما يتعلق بمستوى أداء الأفراد المختبرين، وأسلوب الكتابة، وملاءمة الوقت لعملية التصحيح. لذلك ينبغي العناية بوضع خطة تصحيح الاختبار وتقدير درجاته بغض النظر عن صيغة الاختبار أو محتواه لتحقيق درجة عالية من الموضوعية.

(4) التوازن **Balance**: فلكي يكون الاختبار صادقاً، فإنه يجب أن يقيس السلوك المتعلق بمجال المحتوى المراد قياسه. لذلك فإن التوازن يعني مدى كفاية نسبة عدد المفردات التي تقيس سلوكاً أو ناتجاً معيناً. ويمكن مراعاة ذلك في الاختبارات التحصيلية استناداً إلى جدول مواصفات الاختبار الذي يحدد الأهمية النسبية لكل من المحتوى والأهداف التعليمية المراد قياسها.

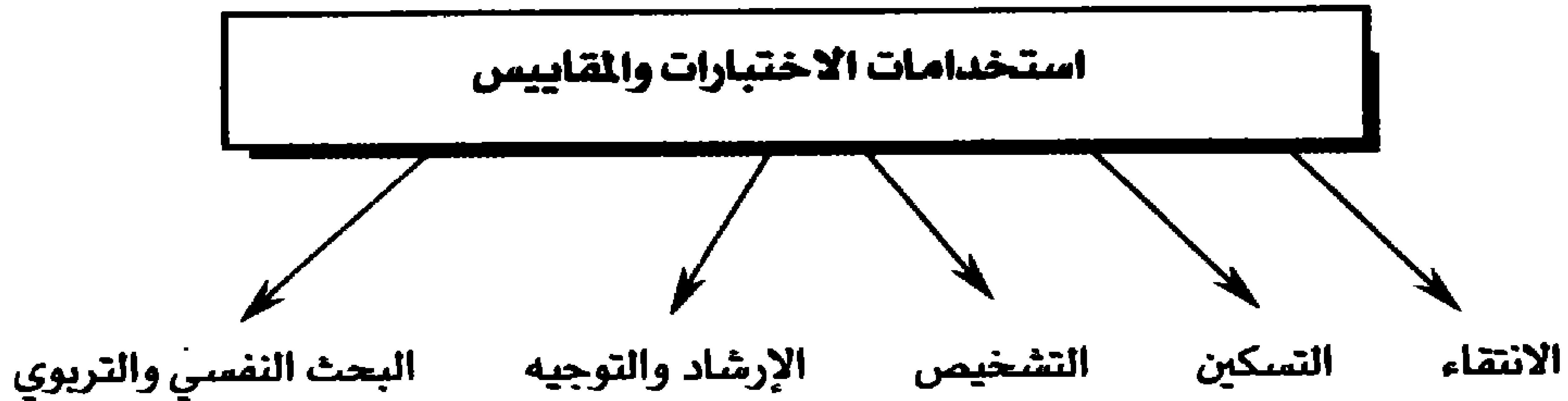
(5) الواقعية **Practicality**: فالاختبارات والمقاييس يجب أن تكون مفيدة لغرض معين. وهذا يتطلب أن يكون الاختبار واقعياً، من حيث: الظروف، والإمكانات المتاحة لبنائه، وتطبيقه، وتصحيحه، وتقرير نتائجه. كما يجب أن يتميز بالفعالية **Efficiency** من حيث عدد الاستجابات التي يمكن أن نحصل عليها في زمن معين. فمفردات الاختيار من متعدد، أو الصواب أو الخطأ مثلاً يمكن الاجابة على عدد كبير منها في زمن مناسب، كما يمكن تصحيحها آلياً. لذلك فإن الاختبار الذي يشتمل على مثل هذه المفردات يكون أكثر فعالية لو اشتمل على أسئلة تستغرق إجابتها وقتاً طويلاً، مثل أسئلة المقال. غير إنه يجب مراعاة أن كلاً من نوعي الأسئلة له استخداماته المناسبة، ولا يمكن إحلال أي من النوعين محل الآخر.



(6) العدالة Fairness : فالاختبارات وبخاصة التي تقيس التحصيل الدراسي ينبغي أن تؤكد المعارف والمهارات والقدرات التي اهتم المعلم بتنميتها لدى طلابه أثناء دراستهم مقرر معين. ولتحقيق العدالة ينبغي الابتعاد عن الأسئلة أو المفردات الاختبارية الخادعة أو الغامضة، كأن تكون صياغتها غير واضحة أو غير مناسبة للمرحلة العمرية للطلاب، أو تحتوي على الفاظ غير مألوفة لهم، أو توحي بإجابات أو تفسيرات غير مطلوبة. فمثل هذه الأسئلة أو المفردات ربما تعوق الطالب المستذكر عن الاجابة الصحيحة. وبذلك تقل درجته التي يحصل عليها عن درجته التي تُعبر عن تحصيله الفعلي، وهو ما لا يجب أن يكون.

استخدامات الاختبارات والمقاييس النفسية والتربوية:

تُستخدم الاختبارات والمقاييس كما أوضحنا في اتخاذ قرارات متعددة تتعلق بالأفراد بأقل درجة من المخاطرة، فلكل اختبار أو مقياس استخدامات تصلح لأغراض معينة ولا تصلح لأغراض اخرى. لذلك فإن انتقاء أداة القياس المناسبة للغرض المرجو يزيد احتمال اتخاذ قرار أفضل. ولا يعتمد ذلك فقط على جودة الأداة (الاختبار أو المقياس)، وإنما على اعتبارات أخرى، مثل: مدى كفاية تحديد مكونات السمة المراد قياسها، ومدى مواءمتها في قياس التباين الفعلي بين الأفراد، وكيفية استخدام البيانات أو الدرجات المستمدة منها. ويوضح الشكل التخطيطي (1 - 6) التالي بعض أهم استخدامات الاختبارات والمقاييس التربوية والنفسية للمعاونة في اتخاذ القرارات المختلفة:



شكل (1 - 6): يوضح بعض أهم استخدامات الاختبارات والمقاييس التربوية والنفسية

وفيما يلي توضيحاً لكل هذه الاستخدامات:

الانتقاء Selection: فالاختبارات والمقاييس يمكن أن تُستخدم في التنبؤ بالنجاح أو الفشل في الأداء الأكاديمي أو المهني. ففي مواقف الانتقاء يتقدم عادة عدد من الأفراد يزيد عما هو مطلوب، لذلك ينبغي اتخاذ قرار يتعلق بقبول أصلح الأفراد للالتحاق بدراسة أو مهنة معينة، ويكون الغرض من الاختبار المعاونة في اتخاذ قرار انتقاء الأفراد الذين يكون احتمال نجاحهم كبيراً.

وعندما تُستخدم الاختبارات والمقاييس في الانتقاء، ينبغي أن تكون هناك علاقة بين درجاتها والنجاح في البرنامج التعليمي أو العمل. فإذا لم تبين الدراسات أن الاختبارات تميز بين الأفراد المحتمل نجاحهم، والذين لا يحتمل نجاحهم، فإن هذه الاختبارات تكون في حاجة إلى تعديل أو استبعاد. غير أن قرارات الانتقاء لا تكون عادة صادقة تماماً، لذلك يكون هناك دائماً احتمال قبول أفراد يتبين فيما بعد إخفاقهم، أو عدم قبول أفراد كان من الممكن أن ينجحوا اذا تم قبولهم. وينبغي التقليل من هذين الخطأين قدر الامكان. ويمكن تحسين قرارات الانتقاء بجعل المهام، أو المهارات، أو القدرات التي يشتمل عليها الاختبار مناظرة لمتطلبات النجاح في الدراسة أو المهنة المعينة.

التسكين Placement: بعد انتقاء الأفراد لدراسة أو مهنة معينة ينبغي تسكينهم في أحد البرامج أو الأقسام المناسبة. والغرض من ذلك إحداث مزاوجة أمثل بين سمات الفرد التي تم قياسها، ومتطلبات الدراسة أو العمل.

والاختبار الذي يصلح للانتقاء ربما لا يصلح للتسكين، فاختبار الاستعداد العام الاكاديمي مثلاً ربما يفيد في التنبؤ بالنجاح في الدراسة بالكلية، ولكن ليس بالضرورة يكون ذا فائدة في التنبؤ بمن يحتمل نجاحهم في برنامج دراسي معين.

التشخيص Diagnosis: يهدف لمقارنة جوانب القوة والضعف لدى الفرد في مجالات متعددة من أجل تحديد أسباب الصعوبات التي يواجهها أثناء أدائه مهاماً معينة، بحيث يمكن توفير الاساليب المناسبة للعلاج. فمثلاً الاختبار التشخيصي في القراءة يمكن أن يزودنا بدرجات في مجالات مختلفة، مثل: المفردات اللغوية، ومعاني الكلمات والجمل، ومعدل القراءة، وذلك بهدف التحديد الدقيق لجوانب الضعف والقوة في القراءة لدى التلميذ. ومقياس الشخصية ربما يشتمل على مجالات فرعية تتناول قضايا متعلقة بالعلاقات الأسرية والاجتماعية، والاستقرار الانفعالي، وذلك لتحديد أو تشخيص المجال الذي يعاني فيه الفرد من صراعات. ويساعد هذا التشخيص في تقديم برنامج إرشادي أو علاجي.



والاختبارات التحصيلية العادية تعد اختبارات مسحية Survey Tests، حيث تُستخدم في تقدير المستوى العام لأداء الطالب في مادة دراسية معينة. لذلك فإن محتوى الاختبار يكون متسعاً، مما يجعل من الصعب استخدامه في إعطاء صورة صادقة عن جوانب القوة والضعف في تحصيل الطالب. اما الاختبار التشخيصي Diagnostic Test فإنه يقيس التحصيل بدرجة أكثر عمقاً وتفصيلاً نظراً لتركيزه على قياس مكونات المادة الدراسية أو المهارة المعينة. لذلك ينقسم الاختبار التشخيصي عادة إلى عدد من الاختبارات الفرعية، وتقدر درجة لكل طالب في كل اختبار فرعي وهذا يجعل الاختبار أطول نسبياً من الاختبار المسحي.

غير إن بناء المقاييس التشخيصية يحتاج إلى منهجيات وتقنيات متطورة لا تزال في حاجة إلى مزيد من البحث والتطوير.

الإرشاد والتوجيه Counseling & Guidance: تُستخدم الاختبارات والمقاييس في كثير من الأحيان في معاونة الأفراد في اختيار البرامج الدراسية، والمهن، والوظائف الأكثر ملاءمة لهم.

فمعرفة درجات الفرد في عدد من اختبارات الاستعدادات، وبطاريات اختيارات التحصيل، ومقاييس الميول، واستبيانات الشخصية، وسجلات التحصيل الدراسي السابق، والمعلومات التي يتم جمعها من خلال المقابلات، تُمكن المرشد التربوي أو المهني من تلخيص جميع هذه البيانات، والاستناد إليها في معاونة الفرد في عمل تنبؤات تتعلق بمستقبله التعليمي أو المهني. وعادة لا يستخدم اختبار أو مقياس واحد في هذا الشأن، وإنما تطبق سلسلة من الاختبارات كما أشرنا. وتعد قرارات الإرشاد والتوجيه قرارات فردية، على العكس من القرارات المؤسسية التي تتخذها المدارس والجامعات، والهيئات الاجتماعية، والمصانع، وغيرها؛ فقرارات الإرشاد والتوجيه تتعلق بفرد معين، وربما لا تناسب فرداً آخر.

البحث النفسي والتربوي Research: تُستخدم الاختبارات والمقاييس في كثير من الاحيان في التحقق من صحة الفروض؛ فعلماء النفس والتربية يستخدمون الاختبارات في قياس متغيرات تتضمنها فروضهم البحثية من أجل التوصل إلى نظريات تفسر السلوك الإنساني. وبذلك تعد الاختبارات والمقاييس أدوات علمية، ربما لا يكون لها تطبيقات عملية مباشرة في حياة الأفراد.

وبالطبع ليست هذه هي الاستخدامات الوحيدة للاختبارات والمقاييس التربوية والنفسية، اذ

أن هناك العديد من الاستخدامات التي تناسب مختلف الأغراض. ولعل استخدام الاختبارات في المجالات التربوية يعد أكثر الاستخدامات انتشاراً في الدول المختلفة. فالاختبارات تستخدم في تحسين وإثراء تعلم الطلاب، وتقويم تحصيلهم، ومتابعة تقدمهم الدراسي، وتقديم تغذية راجعة للمعلم والطالب. كما تعد الاختبارات مصدراً لزيادة دافعية الطلاب، وتطوير البرامج والمناهج التعليمية، وتقويم مختلف مكونات المنظومة التربوية.

مفهوم التقويم، وعلاقته بالاختبارات والمقاييس Evaluation:

يعد مفهوم التقويم أكثر اتساعاً من كل من مفهومي القياس والاختبارات، فنحن نحاول في القياس تحري الموضوعية قدر الإمكان، وتقليل دور الأحكام إلى أقل حد ممكن، ولكن التقويم يعد عملية تؤكد دور الأحكام استناداً إلى أنواع مختلفة من البيانات الكمية والكيفية التي يتم جمعها وتحليلها، وذلك للتوصل إلى ادلة تفيد في تعرف القيمة، أو الجودة، أو الفاعلية، أو اتخاذ قرارات معينة، أو معاونة فرد آخر في اتخاذ القرار.

فالمرشد التربوي يعاون الطالب في جمع بيانات متنوعة مناسبة عن تحصيله وقدراته وميوله، والتوصل منها إلى معلومات يُتخذ في ضوئها قراراً معيناً يتعلق بالطالب. ففي هذه الحالة تُوجَّه عملية التقويم نحو الاختيار الصائب من بدائل متعددة.

ومن هذا يتضح أن مفهوم التقويم يرتبط ارتباطاً وثيقاً بمفهومي القياس والاختبار. فعلى الرغم من أنه ربما يبدو أن هذه المفاهيم تشير إلى عمليات منفصلة، الا انها متداخلة. فعملية التقويم شأنها شأن أي اختبار، تكون مفيدة إذا أدت إلى بيانات دقيقة ومتسقة وصادقة، بحيث يمكن الاعتماد عليها في إصدار أحكام صائبة، كما أنه من الضروري أن نتحرى في عملية التقويم موضوعية تفسير البيانات، وعدم تأثر الأحكام بالتحيزات الشخصية قدر الإمكان.

وينبغي أن تعتمد عملية التقويم على أنواع مختلفة من أدوات القياس، والربط بين البيانات المستمدة منها لكي يتم فهم السلوك المراد تقويمه فهماً افضل، ومعاونة الأفراد. وهذا يعني أن عملية التقويم لا تقتصر على استخدام الاختبارات فقط، وانما يمكن أن تستخدم قوائم الملاحظة، واساليب المقابلة، واستبيانات الميول والاتجاهات والشخصية، والطرق الإسقاطية، وغيرها.

ومن الجدير بالذكر أن التقويم ليس عملية واحدة، وإنما تتضمن سلسلة من الأنشطة أو عدد من الخطوات. فاتخاذ القرارات التربوية مثلاً يستند إلى الأحكام، التي تستند بدورها إلى المعلومات. والترابط بين مفاهيم القرارات، والأحكام، والمعلومات يعرِّف مفهوم التقويم. لذلك فإن عملية التقويم تمر بمراحل متعددة نوجزها فيما يلي:

(1) مرحلة الاعداد: وتتضمن تحديد نوع الأحكام والقرارات المراد اتخاذها، ووصف المعلومات المطلوبة، وكيفية الحصول عليها.

(2) مرحلة جمع البيانات: وتتضمن الحصول بالفعل على المعلومات، وتحليلها وتدوينها.

(3) مرحلة الحكم واتخاذ القرار: وتتضمن الموازنة بين المعلومات التي تم التوصل إليها، ومحك معين لإصدار الأحكام، ومن ثم اتخاذ قرار، أي الاختيار من بين بدائل افعال. فنظام اتخاذ القرارات ينبغي أن يلقى الضوء على أوجه التماثل والاختلاف بين مختلف البدائل، بحيث يمكن اختيار البديل المناسب بسهولة. وينبغي بعد ذلك تلخيص النتائج وكتابة، تقرير التقويم الذي يُوجه لفئة معينة، مثل: الآباء، أو المعلمين، أو الإداريين، أو غيرهم من المعنيين بنتائج التقويم.

أخلاقيات استخدام الاختبارات:

على الرغم من أن الاختبارات والمقاييس لها فوائد واستخدامات جيدة متعددة أوضحناها فيما سبق، إلا أن إساءة استخدام هذه الأدوات يكون له عواقب سلبية كثيرة. وهذا يتطلب الإجابة على تساؤلات عديدة (Mehrens & Lehman, 1969) مثل:

(1) ما مستوى تأهيل الشخص الذي يوكل إليه تطبيق الاختبارات وتفسير نتائجها؟

(2) ما نوع الاختبارات التي ينبغي استخدامها؟

(3) كيف تُستخدم نتائج الاختبارات؟

(4) من يسمح له بالاطلاع على نتائج الاختبارات؟

(5) هل الاختبارات تتميز بالعدالة؟

(6) هل عدد الاختبارات التي تطبق على الطلاب كثيرة أم قليلة؟

وهذه التساؤلات مترابطة، حيث إن أنواع الاختبارات التي ينبغي تطبيقها يعتمد على مستوى تأهيل مستخدميها، وأغراض تطبيق هذه الاختبارات.

وقد حرصت الجمعية الامريكية لعلم النفس (APA) على إصدار دليل متجدد بعنوان "معايير للاختبارات التربوية والنفسية، وكذلك دليل "المعايير الأخلاقية للمشتغلين بعلم النفس"، ويشتمل الدليلان على أقسام متعددة تقدم خطوطاً إرشادية لتفسير درجات الاختبارات، وخصوصياتها، وسريتها. وكذلك معايير تتعلق باستخدام الاختبارات، ومبادىء التحقق من صدقها واستخدامها في انتقاء الأفراد، وغير ذلك من الجوانب الأخلاقية والاجتماعية للاختبارات واستخداماتها. ويمكن للقارىء أو الباحث المهتم بذلك الرجوع إلى هذين الدليلين.

غير أنني سوف اقتصر على أيجاز بعض أوجه النقد والجدل حول الاختبارات والمقاييس النفسية والتربوية، وإساءة استخدام وتفسير نتائج الاختبارات المقننة، وبعض الإرشادات للتغلب على ذلك. فقد أشارت ادبيات القياس والتقويم إلى أوجه النقد التالية:

(1) الاختباؤات تكون متحيزة ضد بعض الأفراد في مواقف معينة، وبخاصة الاقليات والمحرومين.

(2) التنبؤ باستخدام الاختبارات يكون منقوصاً، ويتطلب الاستعانة بأنواع أخرى من المعلومات، كما أن المحكات المستخدمة في التحقق من صدق الاختبار ليست مثالية.

(3) الاختبارات تؤدي إلى تصنيف الأفراد تصنيفات قاطعة، مما يعني عدم الاعتراف بأن الأفراد دائمي التغير، وأن نتائج الاختبارات تعد بمثابة قياسات تقريبية لسمات الأفراد وخصائصهم.

(4) الاختبارات تفترض قياس سمات أو خصائص موروثة.

(5) نتائج الاختبارات تسهم بدرجة أساسية في تغيير نظرة الآخرين عن الفرد وفقاً لما يتوقعونه منهم.

(6) الاختبارات تؤثر تأثيراً غير مرغوب فيه في نمط تفكير الفرد وفقاً لنوع مفرداتها.

(7) الاختبارات تسهم عن طريق إساءة استخدامها في تشكيل المناهج التعليمية، وتعوق التطوير التربوي.

(8) الاختبارات تؤثر تأثيراً سلبياً في مفهوم الذات لدى الفرد، ومستوى طموحه، وبخاصة إذا قام بتطبيقها وتفسير نتائجها أخصائي نفسي غير متمرس.

(9) الاختبارات تتدخل في خصوصيات الأفراد، وبخاصة الاختبارات والمقاييس المتعلقة بالجوانب الوجدانية والشخصية. وهذا يؤثر في عملية انتقاء الأفراد لوظائف معينة، إذ ربما يؤدي ذلك إلى استبعاد أفراد أكفاء من أعمال يصلحون لها استناداً إلى نتائج هذه الاختبارات التي يرى النقاد أنها تستكشف عقول المتقدمين للوظائف.

وقد حاول بعض علمـــاء النفس وخبراء القياس تفنيد هذه المزاعـــم، ورأوا أن كثيراً منها لا يرجع إلى الاختبارات والمقاييس في حد ذاتها، وإنما إلى إساءة استخدامها بما لا يتفق مع الأهداف المرجوة منها. فنوعية الاستخدام في مختلف مجالات الخدمات النفسية والتربوية، وكفاءة القائمين بذلك، يزيد من فاعلية أدوات القياس التي يعتني بتصميمها واعدادها.

لذلك حددت الجمعية الامريكية لعلم النفس 19 مبدءاً أخلاقياً ينبغي الاستناد إليه في استخدام الاختبارات والمقاييس. وتتناول هذه المبادىء: الكفأيات اللازم توافرها فيمن يستخدم الاختبارات، وأدوات القياس وإجراءاتها، وحمأية خصوصيات الأفراد، وسرية الاختبارات والبيانات المستمدة مـــنها، وتوصيل النتائج للمختبرين، وحقـوق الأقلـيات والمحرومين. وسوف أوضح فيما يلي بعض هذه المبادىء بأيجاز (Aiken, 1976, Anastasi, 1976; Sax, 1974):

الخصوصية Confidentiality: فدرجات الاختبارات والتواصل اللفظي بين المرشد التربوي أو الأخصائي النفسي والفرد المختبر تعد معلومات لها خصوصيتها. ولا يجوز نقل هذه المعلومات إلى الآخرين دون موافقة الفرد، إلا إذا كان هناك ضرر واضح يمكن أن يعود على الآخرين أو على المجتمع. وعلى الرغم من أن الدرجات التي يتم الحصول عليها في البرامج الاختبارية التي تجريها مؤسسات معينة تكون أقل سرية منها في حالات الإرشاد الفردي، إلا أنه ينبغي في الحالتين تحديد الجهة التي تحتاج إلى معرفة نتائج الاختبارات والمقاييس، وأنسب طرق توصيل المعلومات المتعلقة بها.

السرية Test Security: فالاختبارات تعد أدوات قياس يستخدمها المشتغلون بعلم النفس، والأخصائيون، لذلك فإن توزيع هذه الاختبارات ينبغي أن يقتصر على أولئك الذين يتمتعون بكفاءات فنية تمكنهم من استخدام هذه الأدوات استخداماً مناسباً. فمعظم الاختبارات المقننة ينبغي ألا يسمح بتداولها الا للمتخصصين، وهذا يتطلب تنظيم عملية نشرها وتوزيعها تجنباً للأضرار التي تنجم عن إساءة استخدامها.

تطبيق الاختبارات Test Administration : عندما تُطبق الاختبارات والمقاييس التربوية والنفسية ينبغي توضيح الغرض من الاختبار لكل فرد مختبر، وفيم سوف تُستخدم نتائجه. كما ينبغي إخبار الأفراد بحقوقهم المتعلقة بالخصوصية، وانه يمكنهم رفض الإجابة على الأسئلة التي يرون أنها غير ملائمة أو تتدخل مباشرة في خصوصياتهم.

كفايات مستخدمي الاختبارات User Qualifications: يمكن التقليل من إساءة استخدام الاختبارات بدرجة كبيرة إذا كان القائمون بتطبيق الاختبارات وتفسير نتائجها لديهم الكفايات المناسبة لذلك. وتختلف هذه الكفايات باختلاف نوع الاختبار أو المقياس، فالاستخدام المناسب للاختبارات الفردية في الذكاء، ومعظم مقاييس الشخصية، يحتاج إلى تدريب مكثف لمدة طويلة بإشراف متخصصين. أما الاختبارات التربوية والمهنية، فإنها تحتاج إلى تدريب متخصص أقل، بحيث يمكن اختيار الاختبارات المناسبة للغرض من القياس وطبيعة الأفراد المختبرين، وكذلك تقييم الخصائص الفنية لهذه الاختبارات، مثل: المعايير، والصدق، والثبات. ويكون المختبر مدركاً للظروف التي تطبق فيها الاختبارات والتي ربما تؤثر في أداء الأفراد، وتفسير نتائج الاختبارات باستخدام الدرجات الناتجة عنها في ضوء معلومات أخرى تتعلق بالفرد، مع مراعاة دقة الاستنتاجات في تفسير الدرجات.

الفصل الثاني

الإحصاء في القياس النفسي والتربوي

- مقدمة
- أنواع البيانات الاختبارية
- تنظيم الدرجات
- مقاييس النزعة المركزية
- العلاقة بين مقاييس النزعة المركزية
- مقاييس التشتت
- التوزيع الاعتدالي للدرجات
- مقاييس العلاقة
- الارتباط والتنبؤ



يعتمد تحديد الخصائص السيكومترية للاختبارات والمقاييس النفسية والتربوية، وكذلك تفسير درجاتها على بعض الاساليب الاحصائية التي ينبغي استيعابها وتوظيفها توظيفاً مستنيراً في التعامل مع الدرجات المستمدة منها. فالفائدة المرجوة من نتائج القياس تعتمد بدرجة كبيرة على دقة تحليل هذه النتائج بفكر منفتح. ويرجع ذلك إلى ان الدرجات المستمدة من أدوات القياس لا تعني شيئاً ما لم يجرى عليها بعض التحليلات الاحصائية المناسبة لكي يتم تفسيرها والافادة منها، واجراء مقارنات بين الافراد او المجموعات في سمات او خصائص معينة. فالتحليل الاحصائي يعد اذن اداة تساعد في فهم هذه السمات او الخصائص استناداً إلى الدرجات او القيم العددية المستمدة من الاختبارات والمقاييس. لذلك سوف يتناول هذا الفصل المفاهيم والمبادىء والاساليب الاحصائية الاساسية التي تستخدم في وصف توزيعات درجات الاختبارات والمقاييس النفسية والتربوية من حيث: النزعة المركزية، والتشتت، والعلاقات القائمة بين الدرجات. وهذا يساعد في تحليل نتائج الاختبارات التي يعدها الباحث بنفسه، وفهم البيانات الاساسية التي تزودنا بها أدلة الاختبارات المنشورة، والبحوث التي تُجرى عليها.

سبق أن عرَّفنا القياس بأنه تعيين اعداد تناظر السمات الانسانية طبقاً لقواعد محددة. وهذه الاعداد هي الدرجات المستمدة من الاختبارات والمقاييس. وتختلف هذه الدرجات باختلاف نوع اداة القياس المستخدمة، فقد تكون درجات مفردات اختيار من متعدد او مقال، او عدد افعال معينة يقوم بها الفرد او تكرارها. ويمكن ان نطلق على هذه الدرجات او القيم العددية اسم القيم الملاحظة، او البيانات الاختبارية، او البيانات الخام، او قيم المتغير الذي اجرى قياسه. وهذه البيانات هي التي يتم معالجتها إحصائياً.

كمية Quantitative، وكيفية او نوعية Qualitative وفقاً لمستوى قياس المتغير او المتغيرات التي يتم قياسها لدى الأفراد. وقد سبق ان اوضحنا المستويات بالتفصيل في الفصل الأول.

ويوضح الشكل التخطيطي (2 - 1) التالي علاقة نوعي المتغيرات بمستويات القياس:

البيانات الاختبارية

متغيرات كمية			متغيرات كيفية (نوعية)
متغيرات مُتصلة			متغيرات منفصلة
رتبية	فترية	نسبية	اسمية

: يوضح علاقة المتغيرات الكمية والكيفية بمستويات القياس

ويتضح من شكل (2 - 1) ان المتغيرات الكمية التي يكون التغير فيها تغيراً في المقدار تكون متصلة، اي تختلف بمقادير صغيرة صغراً لا نهائياً، ومثال ذلك: الطول، والعمر، والوزن، ودرجات الاختبارات التحصيلية والعقلية. أما المتغيرات الكيفية فهي التي تستخدم في تصنيف الافراد في سمة معينة في اقسام متنافية مثنى مثنى، لذلك تعد متغيرات منفصلة. ومستوى قياس هذه المتغيرات يكون اسمياً، في حين ان مستوى قياس المتغيرات الكمية يكون رتبياً، أو فترياً، او نسبياً. وتختلف الأساليب الاحصائية التي تعالج كلا من هذين النوعين من البيانات. غير اننا سوف نركز على البيانات الكمية التي نحصل عليها من مختلف الاختبارات والمقاييس النفسية والتربوية.



إن الدرجات الخام المستمدة من الاختبارات يصعب تفسيرها لأنها غير مرتبة أو منظمة. لذلك فإن تجميع هذه الدرجات في توزيع تكراري يعد ضرورياً لمعرفة عدد الأفراد الذين حصلوا على درجات معينة. وهذا يتطلب تقسيم الدرجات إلى مجموعات او فئات حدودها معلومة، ومن ثم تعيين كل درجة في الفئة المناسبة. وينبغي ان يكون عدد الفئات معقولاً مثل 10 أو أي فئة لكي يمكن التعامل بسهولة مع جداول التوزيعات التكرارية للدرجات.

لكي نكون جدول التوزيع التكراري لمجموعة من الدرجات يمكن اتباع الخطوات التالية:

وهو الفرق بين أصغر درجة وأكبر درجة، وإضافة واحد صحيح إلى هذا الفرق للحصول على المدى الفعلي.

ولتوضيح ذلك نفترض أن لدينا درجات اختبار تحصيلي لمجموعة من الطلاب كالتالي:

118, 118, 118, 121, 109, 107, 123, 131, 131, 126 , 113
108, 109, 110, 110, 111, 112, 113, 114, 116, 116, 118
102 , 104 , 103 , 97 , 127 , 139 , 94 , 94 , 96 , 101, 105
112, 114, 114, 116, 117, 119, 120, 124, 129, 129, 128
121, 122, 131, 105, 109

ونلاحظ من هذه الدرجات أن أقل درجة (94) ، وأعلى درجة (139).

وبذلك يكون مدى الدرجات (139 - 94) + 1 = 45 + 1 = 46

نقسم المدى إلى اقسام تسمى الفئات، وتستخدم في تجميع الدرجات. فإذا قررنا ان يكون عدد الفئات (10)، فإن سعة الفئة المناسبة في هذه الحالة = 46 ÷ 10 = 4.6 أي (5) تقريباً.

يحتوي هذا التوزيع على مجموعة من الفئات المرتبة

بحيث تشمل جميع الدرجات. ونبدأ بأقل الدرجات لتكوين الفئة الدنيا (94)، ولكن يفضل اختيار الدرجة الأقل من (94) وتقبل القسمة على سعة الفئة (5) فنجدها (90)، وذلك لتيسير العمليات الحسابية التي يمكن إجراؤها فيما بعد باستخدام هذا الجدول. وبهذا تكون الفئة الدنيا (90 - 94)، وهذه الفئة تبدأ في الحقيقة من (89.5) وتنتهي عند (94.5) وبذلك تكون سعة الفئة بالفعل (94.5 - 89.5 = 5) .

(4) [illegible]: نعين كل درجة من الدرجات السابقة في الفئة المناسبة لها، وذلك في العمود الأول في جدول التوزيع التكراري، ونضع علامة (1) في العمور الثاني كلما تكررت درجة في فئة معينة، ويفضل وضع كل خمس علامات في حزمة واحدة، وذلك بوضع شرطة مائلة تقطع العلامات الأربع، ونضع في العمود الثالث عدد هذه العلامات لكل فئة، كما هو موضح بجدول التوزيع التكراري (2-1) التالي:

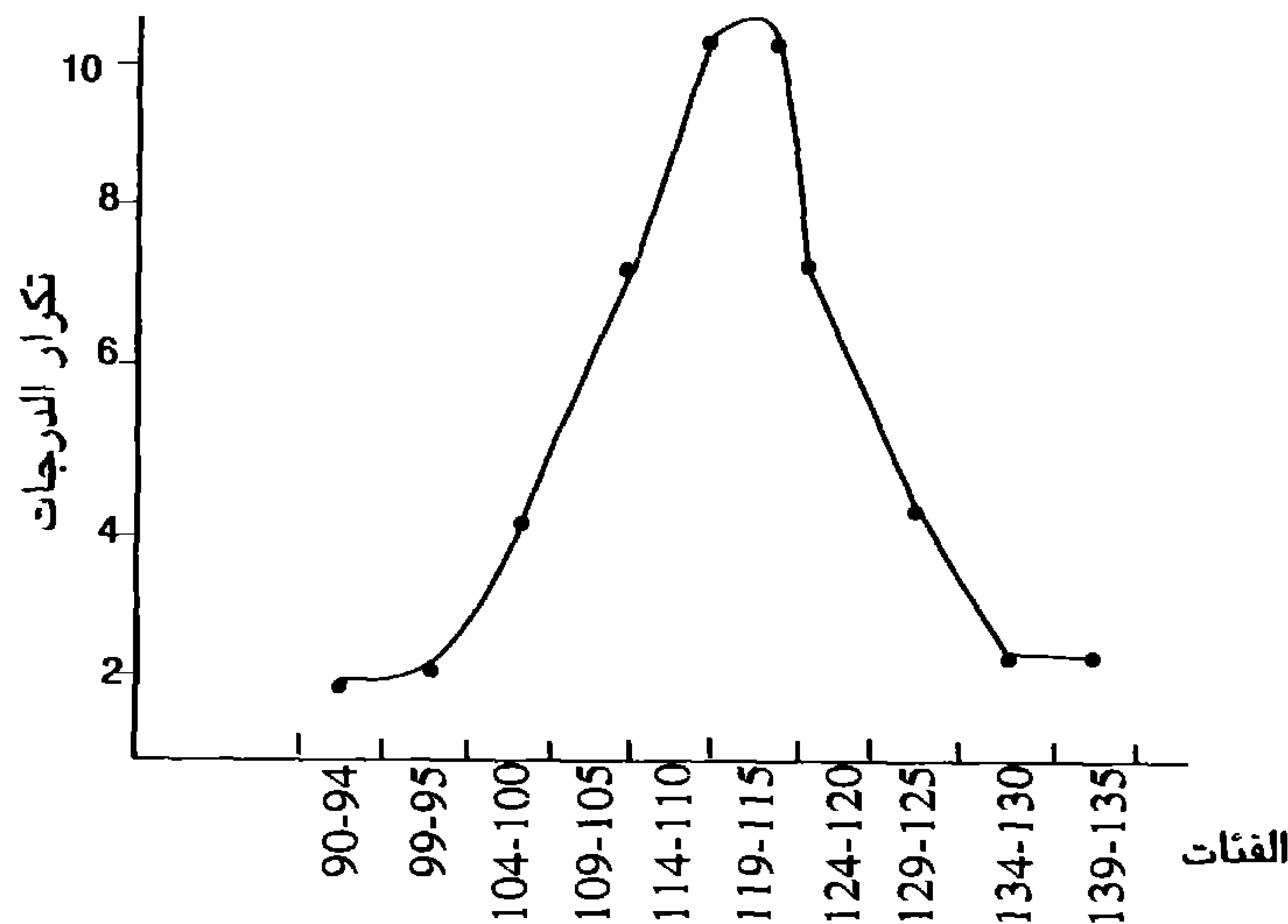
الفئات	علامات الدرجات	تكرار الدرجات
139 - 135	11	2
134 - 130	111	3
129 - 125	1111	4
124 - 120	~~1111~~ 1	6
119 - 115	~~1111~~ ~~1111~~	10
114 - 110	~~1111~~ ~~1111~~	10
109 -105	~~1111~~ 11	7
104 - 100	1111	4
99 - 95	11	2
94 - 90	11	2
ن = 50		

جدول (2 - 1): جدول توزيع تكراري لمجموعة من الدرجات (ن = 50)



وهذا الجدول يساعد في معرفة عدد الأفراد الذين حصلوا على درجات تقع في فئة معينة، دون تحديد دقيق لدرجة كل منهم. فمثلاً نلاحظ ان (6) افراد حصلوا على درجات تقع في الفئة (115 - 119)، ولكننا لا نستطيع معرفة الدرجة الفعلية لكل منهم، او مدى انتشار الدرجات داخل فئة معينة. ونظراً لأن الدرجات تفقد ذاتيتها في هذا الجدول. لذلك فإن افضل تخمين لدرجة اي فرد هو الدرجة التي في منتصف الفئة، أي $\frac{119+115}{2} = \frac{234}{2} = 117$. ومع هذا فإن جداول التوزيعات التكرارية تعد نقطة البدء في اجراء التحليلات الاحصائية للبيانات الاختبارية اذا أجريت هذه التحليلات يدوياً. أما إذا استخدم الحاسوب كما هو المعتاد، فإنه يمكن إدخال الدرجات الخام، ويقوم البرنامج المناسب بتكوين هذا الجدول، وإجراء التحليلات المناسبة.

يمكن تمثيل التوزيع التكراري للدرجات الموضح في جدول (2 - 1) تمثيلاً بيانياً وذلك برسم محورين متعامدين يمثل المحور الأفقي فئات الدرجات، والمحور الرأسي تكرار الدرجات، وتعيين نقط في مستوى المحورين تناظر الفئات والتكرار المقابل لها بمقياس رسم مناسب، والتوصيل بينها بمنحنى يسمى المنحنى التكراري Frequency curve كالتالي:



: المنحنى التكراري لمجموعة الدرجات (ن = 50)

ويمكن باستخدام هذا المنحنى معرفة عدد الأفراد الذين تقع درجاتهم في فئة معينة كما هو الحال في جدول التوزيع التكراري.

نحتاج في بعض الأحيان معرفة عدد الأفراد الذين حصلوا على درجة أقل من درجة معينة لتحديد المركز النسبي لفرد معين بين اقرانه، لذلك يمكن تكوين جدول آخر يسمى جدول التوزيع التكراري المتجمع الصاعد.

ونبدأ في هذا التوزيع بأصغر الفئات، ونجمع تكرارات الفئات الأكبر منها، كما هو موضح بجدول (2 - 2) التالي

[illegible]	[illegible]	[illegible]	[illegible]
139 - 135	2	50	100
134 - 130	3	48	96
129 - 125	4	45	90
124 - 120	6	41	82
119 - 115	10	35	70
114 - 110	10	25	50
109 - 105	7	15	30
104 - 100	4	8	16
99 - 95	2	4	8
94 - 90	2	2	4
ن = 50			

جدول التوزيع التكراري المتجمع الصاعد للدرجات (ن = 50)

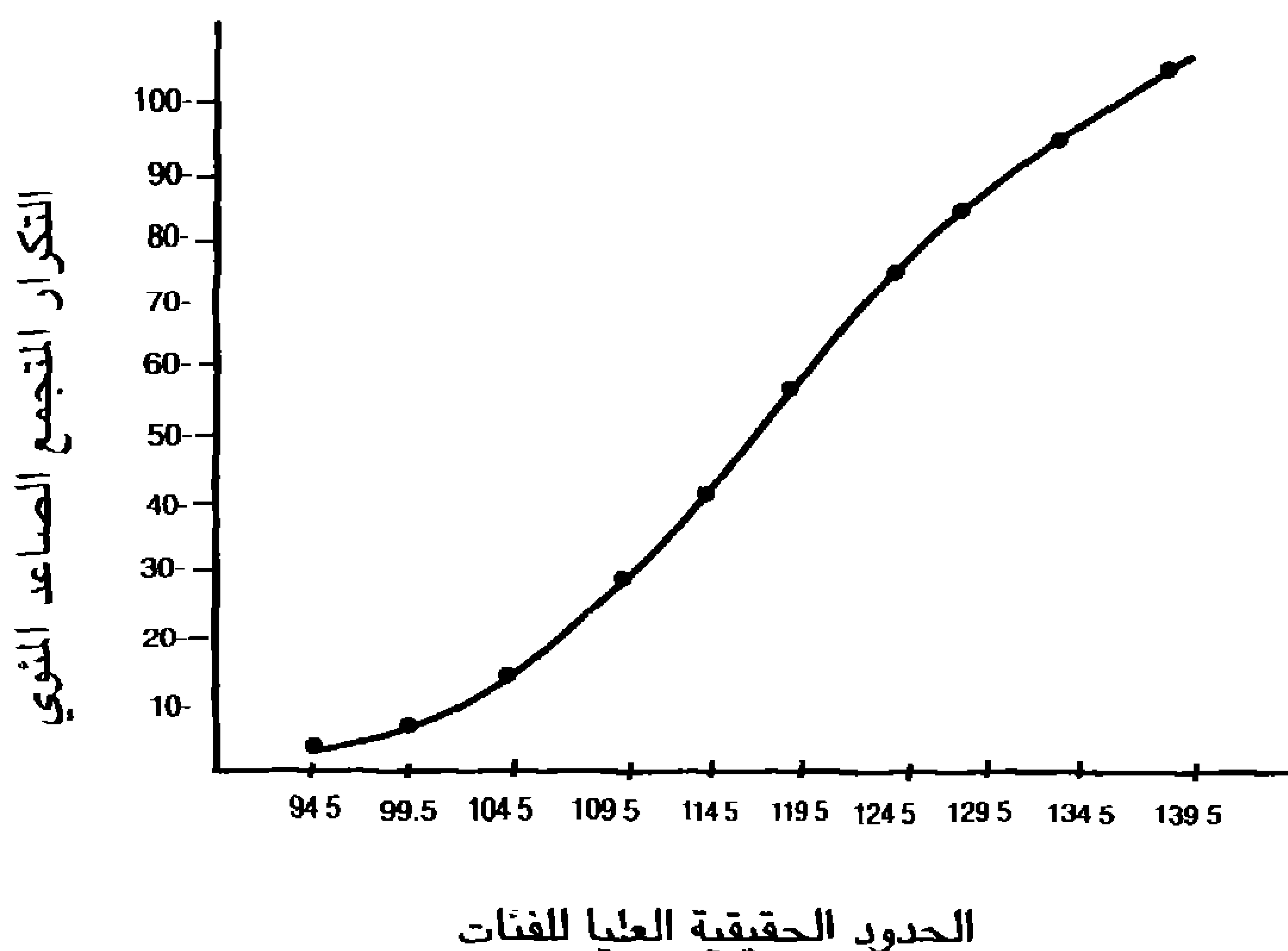


ويلاحظ من هذا الجدول أننا اضفنا عموداً رابعاً يشتمل على التكرار المتجمع المئوي، وذلك بقسمة كل تكرار متجمع صاعد في العمود الثالث على العدد الكلي للأراد (50)، وضرب الناتج × 100. فمثلاً التكرار المتجمع المئوي الذي يناظر التكرار المتجمع الصاعد (15) نحصل عليه كالتالي:

$\frac{15}{50} \times 100 = 30$، وهكذا.

أي ان هناك (15) فرداً (30% من مجموع الطلاب) تقل درجاتهم عن الحد الأعلى الحقيقي للفئة (105-109)، اي تقل عن 109.5.

يمكن تمثيل التوزيع التكراري المتجمع الصاعد بيانياً بنفس طريقة تمثيل التوزيع التكراري. غير ان المحور الافقي في هذه الحالة يمثل الحدود الحقيقية العليا للفئات، والمحور الرأسي يمثل التكرار المتجمع الصاعد او التكرار المتجمع المئوي. ويفضل استخدام التكرار المتجمع المئوي على المحور الرأسي لأهمية ذلك في إيجاد معايير الاختبارات، كما سيتضح في الفصل الرابع.



شكل 2 - 3: المنحنى التكراري المتجمع الصاعد المئوي لمجموعة الدرجات (ن = 50)

ويمكن من شكل (2 - 3) تحديد نسبة عدد الأفراد الذين حصلوا على درجات تزيد عن درجة معينة. فمثلاً اذا اردنا معرفة نسبة عدد الافراد الذين تزيد درجاتهم عن 114.5 فاننا نرسم خطاً رأسياً من النقطة 114.5 التي تقع على المحور الأفقي (الحدود الحقيقية العليا للفئات) حتى يلاقى المنحنى التكراري المئوي في نقطة اخرى نسقط منها عموداً على المحور الرأسي (التكرار المتجمع المئوي)، فتكون نقطة الالتقاء بهذا المحور مناظرة للتكرار المئوي 50، وهذا يعني ان 50% من الأفراد حصلوا على درجات تزيد عن (114.5)، أو أن الفرد الذي حصل على الدرجة (114.5) قد تفوق على 50% من أقرانه.

مقاييس النزعة المركزية:

تفيد هذه المقاييس في معرفة نزعة الدرجات للتمركز حول قيمة معينة. والغرض الأساسي من هذه المقاييس هو التوصل إلى قيمة عددية تمثل بدرجة افضل جميع درجات مجموعة الأفراد في اختبار معين. وبذلك يمكن مقارنة درجة فرد معين بدرجات اقرانه، او مقارنة درجات مجموعتين من الأفراد في الاختبار، وكذلك مقارنة درجات نفس المجموعة في اختبارين مختلفين. وتوجد ثلاثة مقاييس للنزعة المركزية Central Tendency هي: المتوسط، والوسيط، والمنوال.

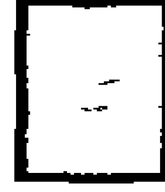
المتوسط الحسابي لمجموعة من الدرجات هو قيمة عددية نحصل عليها بجمع جميع الدرجات وقسمة المجموع على عدد الدرجات. فإذا رمزنا للمتوسط الحسابي بالمرمز (سَ) فإن

فمثلاً متوسط الدرجات 3 ، 5 ، 8 ، 2 ، 7 يساوي $\frac{25}{5} = 5$.

ويتميز المتوسط بأنه يعتمد على جميع الدرجات، ولا تتذبذب قيمته كثيراً من عينة إلى اخرى اذا استمدت العينات عشوائياً من مجتمع معين وكانت ممثلة له. لذلك فإنه يعد اكثر مقاييس النزعة المركزية اتساقاً. كما أنه اذا طُرحت كل درجة في توزيع معين من متوسط الدرجات وجُمعت الفروق الناتجة، فإن هذا المجموع يساوي صفراً. ومستوى القياس المناسب للمتوسط هو المستوى الفتري أو النسبي، حيث يعتمد على جميع الدرجات الفعلية للاختبار.



الوسيط هو النقطة التي تقسم عدد درجات مرتبة في توزيع معين إلى نصفين متساويين. فمثلاً وسيط الدرجات 2 , 5 , 8 , 3 , 7 يتطلب ترتيب الدرجات اما تنازلياً او تصاعدياً كالتالي:



وبذلك تكون قيمة الوسيط = 5، حيث انها تقسم توزيع الدرجات بحيث تزيد درجتان عن الدرجة 5، وتقل عنها درجتان.

واذا كان عدد الدرجات زوجياً فإننا نوجد متوسط الدرجتين اللتين تتوسطان التوزيع.

فإذا أضفنا الدرجة 9 إلى التوزيع السابق بحيث يصبح كالتالي:



فإن قيمة الوسيط = $\frac{7+5}{5} = \frac{12}{2} = 6$.

ويتميز الوسيط بعدم تأثره بالقيم او الدرجات المتطرفة، حيث ان قيمة وسيط الدرجات 1 ، 2 ، 3 ، 4 ، 5 هي نفسها قيمة وسيط الدرجات 1 ، 2 ، 3 ، 4 ، 5 ، 56 وذلك لان قيمته تظل مساوية 3 في الحالتين.

ولكن نظراً لأن الوسيط تتحدد قيمته بالدرجة او الدرجات الوسطى فقط بعد ترتيب مجموعة الدرجات، فإنه اقل اتساقاً من المتوسط، حيث إن قيمته يمكن أن تتغير بدرجة اكبر من عينة إلى اخرى مستمدة من نفس المجتمع. ومستوى القياس المناسب للوسيط هو المستوى الرتبي، حيث يعتمد على ترتيب الافراد في سمة معينة.

المنوال هو الدرجة الأكثر تكراراً في توزيع درجات معين. فمثلاً منوال توزيع الدرجات 2 ، 7 ، 8 ، 8 ، 5 يساوي 8 . وإذا تكررت درجتان اكثر من غيرهما في التوزيع، فعندئذ يكون التوزيع ثنائي المنوال Bimodal. ويتميز المنوال بأنه يمكن حسابه بسهولة، ولكنه يعد اقل

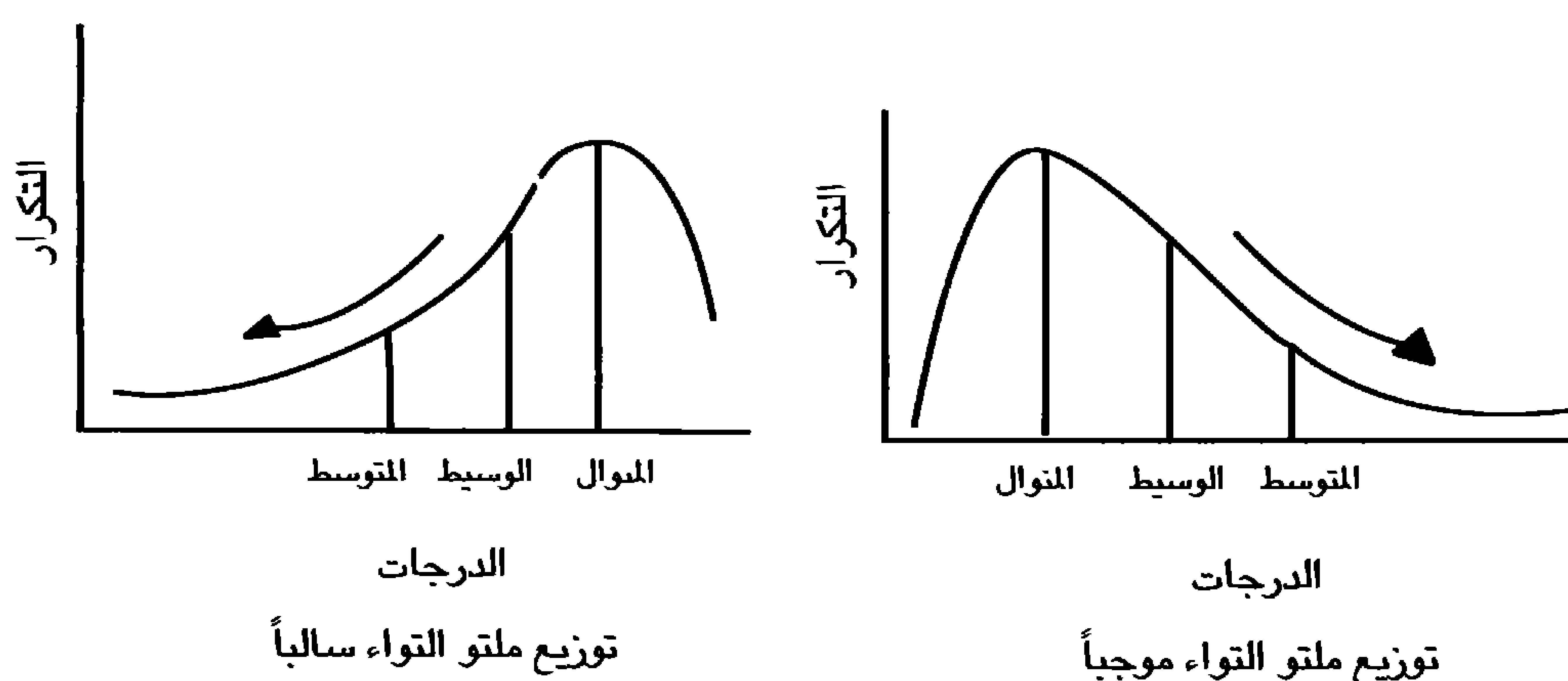
مقاييس النزعة المركزية استقراراً، حيث إن اضافة او حذف درجة واحدة إلى أو من التوزيع تؤدي إلى تغيير قيمة المنوال تغييراً ملحوظاً. ومستوى القياس المناسب للمنوال هو المستوى الاسمي.

تعتمد العلاقة بين المتوسط، والوسيط، والمنوال على شكل توزيع درجات الاختبارات. فإذا كان التوزيع اعتدالياً، او متماثلاً Symmetrical وأحادي المنوال، تكون قيم المقاييس الثلاثة متساوية. أما إذا كان التوزيع ملتوياً التواءً سالباً Negatively Skewed وأحادي المنوال، فإن قيمة المتوسط تكون اقل من قيمة الوسيط، وقيمة الوسيط اقل من قيمة المنوال.

ويحدث العكس إذا كان التوزيع ملتوياً التواءً موجباً Positively Skewed.

لذلك يمكن بمعلومية قيم كل من متوسط ووسيط ومنوال توزيع درجات معين التعرف على اتجاه التواء التوزيع، وذلك بالمقارنة بين القيم الثلاث. ودرجة الالتواء Skewness ربما توضح شيئاً عن المستوى العام لصعوبة الاختبار بالنسبة لمجموعة معينة من الافراد.

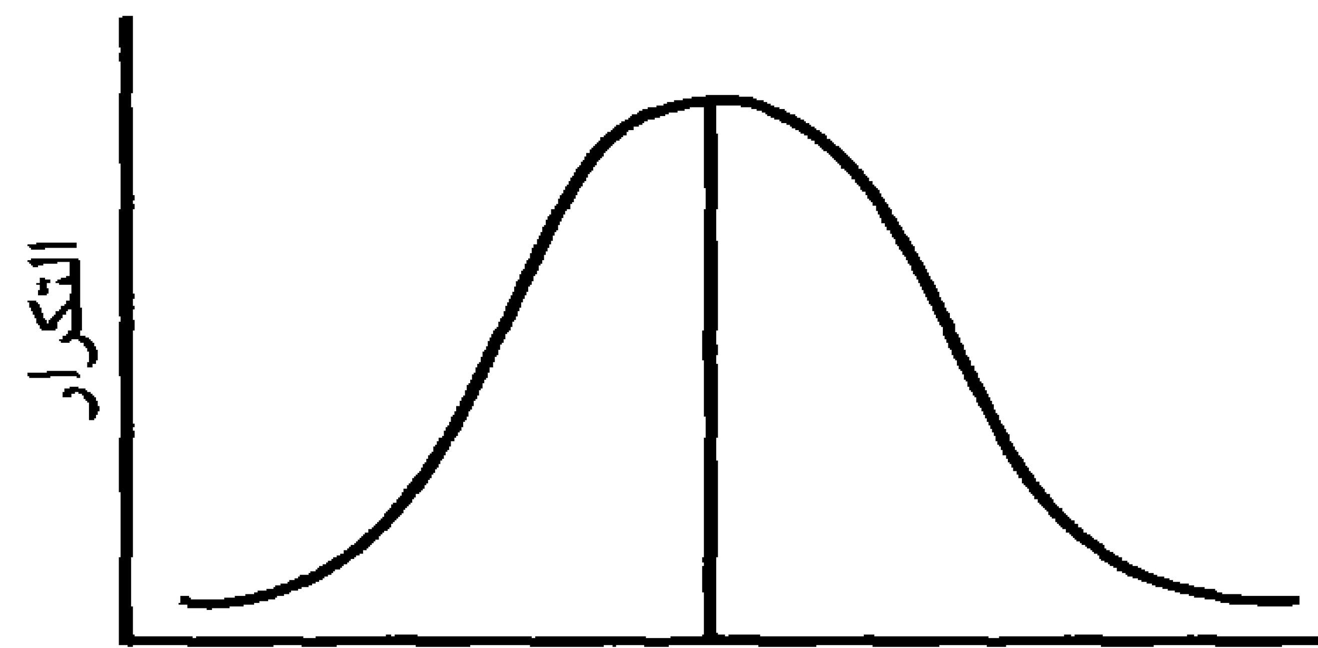
ويوضح الشكل التخطيطي (2 - 4) التالي العلاقة بين المقاييس الثلاثة في كل من هذه الحالات:



توزيع ملتو التواء سالباً

توزيع ملتو التواء موجباً





توزيع ملتو التواء موجباً

شكل 2 : 4 يوضح العلاقة بين المتوسط والوسيط والمنوال.

وتُستخدم في كثير من الأحيان مقاييس النزعة المركزية الثلاثة معاً، وذلك لإلقاء الضوء على شكل التوزيع وتعرف اتجاه التوائه، وهذا يفيد في وصف التوزيع من حيث: النزعة المركزية، والالتواء.

مقاييس التشتت:

يحتاج وصف توزيعات الدرجات أيضاً معرفة تشتت الدرجات او مدى انتشارها. فهذه التوزيعات لا تختلف فقط في النزعة المركزية والالتواء، وانما ايضاً في التشتت Dispersion: فالاعتماد على مقاييس النزعة المركزية وحدها لا يؤدي إلى تفسيرات دقيقة او ذات معنى للبيانات، حيث إن هذه التفسيرات تتطلب ايضاً معرفة مدى التشتت او انتشار الدرجات.

ولتوضيح ذلك نفترض ان مجموعة (أ) من الدرجات كانت كالتالي: 1 ، 2 ، 3 ، 4 ، 5 ومجموعة (ب) : 3 ، 3 ، 3 ، 3، 3 : فهنا نلاحظ ان متوسط كلاً من المجموعتين 3، مما يجعلنا نفترض ان المجموعتين متماثلتين. ولكن بالنظر إلى انتشار درجات كل منهما نجد ان المجموعة (أ) اكثر تشتتاً من المجموعة (ب)، مما يدل على ان المجموعتين ليستا متماثلتين. وهذا يدل على ان مقاييس النزعة المركزية بمفردها تصف توزيعات الدرجات وصفاً ناقصاً، واستكمال الوصف يتطلب معرفة تشتت هذه التوزيعات جزئياً. وتوجد مقاييس متعددة للتشتت نذكر منها ما يلي:

Range المدى

يعد أبسط مقاييس التشتت، حيث يمكن حسابه بإيجاد الفرق بين اكبر الدرجات واصغرها كما فعلنا عند تكوين جدول التوزيع التكراري للدرجات.

فمدى الدرجات: 1 ، 2 ، 3 ، 4 ، 5 يساوي (5 - 4 = 1). وعلى الرغم من أن ايجاد المدى يفيد احياناً في إلقاء نظرة سريعة على انتشار او تشتت الدرجات، الا انه يعتمد على القيمتين المتطرفتين للتوزيع، مما يقلل من الثقة في قيمته لعدم اتساقها. فتغيير إحدى هاتين القيمتين يؤدي إلى تغيير قيمة "المدى" تغييراً جوهرياً.

يعد الانحراف المعياري أكثر مقاييس التشتت استخداماً، حيث يعتمد في حسابه على ايجاد انحراف كل درجة من درجات توزيع معين عن متوسط التوزيع. فكلما زاد هذا الانحراف زادت قيمة الانحراف المعياري. واذا كانت جميع الدرجات متساوية تصبح قيمة الانحراف المعياري صفراً.

ونظراً لاعتماد الانحراف المعياري على جميع الدرجات وليس على الدرجتين المتطرفتين، فإنه أكثر تمثيلاً للتشتت من المدى، مما يزيد من الثقة في قيمه.

ونظراً لأن مجموع انحرافات الدرجات عن متوسط التوزيع تكون مساوية للصفر دائماً، وذلك لأن بعض الانحرافات يكون موجباً والبعض الاخر يكون سالباً، فإنه يفضل تربيع كل من هذه الانحرافات، وايجاد متوسط مجموعها. ويعبر الناتج عن تباين الدرجات Variance، الذي يعد مفهوماً مهماً في القياس النفسي والتربوي، كما سيتضح في الفصول التالية.

والجذر التربيعي للتباين يسمى الانحراف المعياري

حيث ترمز (س) إلى الدرجات.

، (س) إلى متوسط الدرجات.

، (ن) إلى عدد الدرجات



والانحراف المعياري = $\sqrt{\frac{\text{مجـ (س - سَ)}^2}{\text{ن}}}$

ويمكن حساب قيمة كل من التباين والانحراف المعياري للدرجات 1 ، 2 ، 3 ، 4 ، 5 بتكوين جدول (2 - 3) كالتالي:

[illegible]	[illegible]	[illegible]
1	2- = 3 - 1	4
2	1- = 3 - 2	1
3	0 = 3 - 3	0
4	1+ = 3 - 4	1
5	2+ = 3 - 5	4
المجموع 15	صفر	10
المتوسط 3		

: خطوات حساب قيمة التباين، والانحراف المعياري لمجموعة من الدرجات

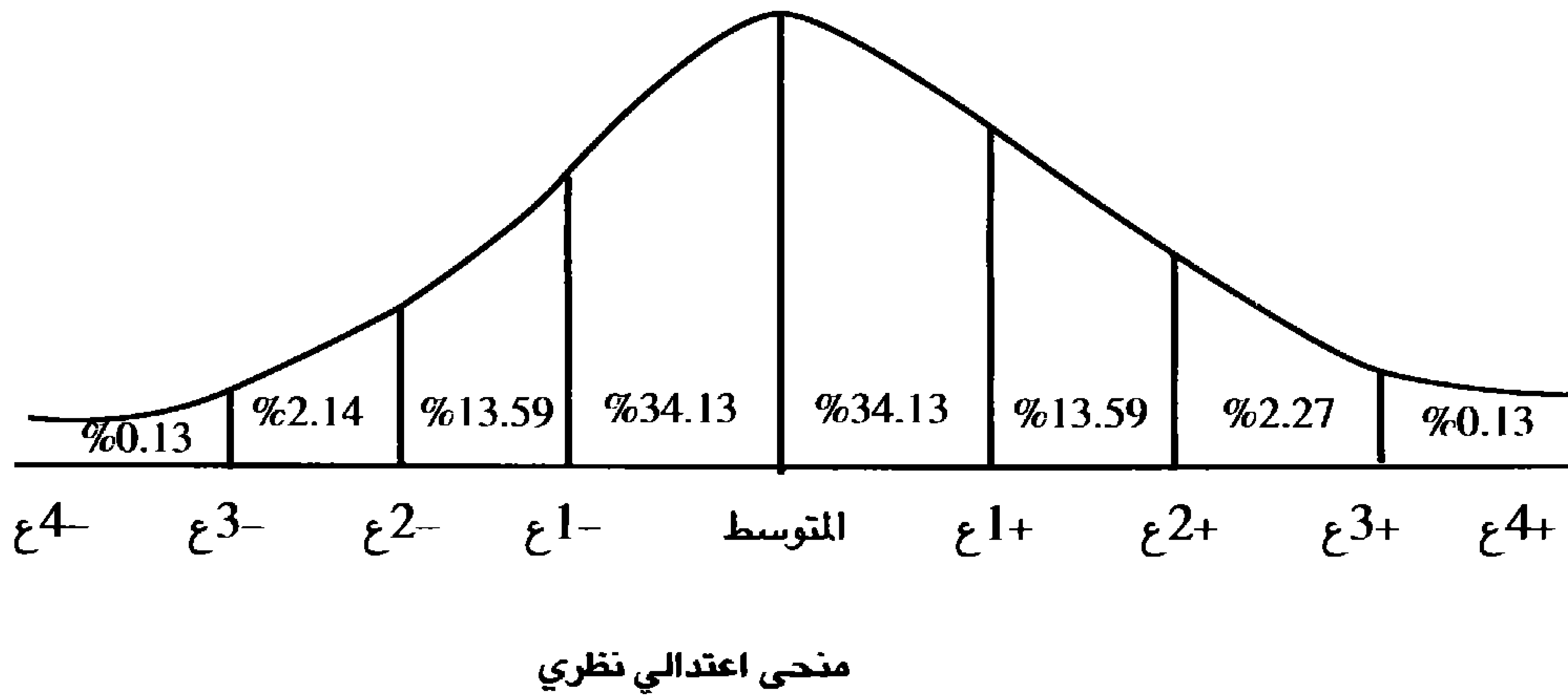
التباين $(\text{ع}^2) = \frac{\text{مجـ ح}^2}{\text{ن}} = \frac{\text{مجـ (س - سَ)}^2}{\text{ن}} = \frac{10}{5} = 2.$

الانحراف المعياري (ع) $= \sqrt{\frac{\text{مجـ ح}^2}{\text{ن}}} = \sqrt{\frac{\text{مجـ (س - سَ)}^2}{\text{ن}}} = \sqrt{\frac{10}{5}} = \sqrt{2} = 1.414$

والانحراف المعياري يعد معياراً لتحديد الانحرافات أو الفروق بين المتوسط وكل درجة من درجات التوزيع. فهو يعد مقياساً احصائياً تزداد قيمته بزيادة تشتت الدرجات، وتكون قيمته صفراً عندما تكون جميع درجات التوزيع متساوية. والقيمة 1.414 التي حصلنا عليها ليست قيمة مطلقة، وإنما هي قيمة نسبية يمكن باستخدامها المقارنة بين تشتت هذا التوزيع، وتشتت

توزيعات درجات اخرى. فإذا كان الانحراف المعياري لثلاثة توزيعات: 1.414، 2.18 ، 3.29 فإن التوزيع الأول للدرجات يكون أقلها تشتتاً، والتوزيع الثالث أكثرها تشتتاً.

يعد التوزيع الاعتدالي لدرجات الاختبارات من التطبيقات المفيدة للانحراف المعياري. فهذا التوزيع متماثل حول المحور الرأسي المار بمتوسط، ويشبه الجرس، وله معادلة رياضية. وعندمانختبر عينة عشوائية من الأفراد ممثلة لمجتمع معين في سمة أو مهارة معينة، فإنه يفترض أن توزيع الدرجات يقترب من التوزيع الاعتدالي، حيث نتوقع ان يحصل عدد قليل الافراد على درجات مرتفعة جداً، او منخفضة جداً، في حين ان معظم الدرجات يميل إلى التجمع بالقرب من الوسط. ولهذا التوزيع خصائص معينة؛ فالمنحنى الممثل له يمتد من كلتا الجهتين إلى اللانهاية، اي ان المنحنى يقترب تدريجياً من المحور الافقي ولكنه لا يمسه مهما مددناه من كلتا الجهتين، وقاعدة المنحنى مقسمة إلى وحدات (انحرافات) معيارية تتراوح من الناحية العملية بين (-4) ، (+4) وحدات. وهناك نسبة مئوية محددة من الافراد تقع في كل منطقة تحت المنحنى، كما هو موضح بالشكل (2 - 5) التالي:



منحنى اعتدالي نظري

ويتضح من شكل (2 - 5) ان:

نسبة 68% تقريباً من الأفراد تقع بين ±1 انحراف معياري.



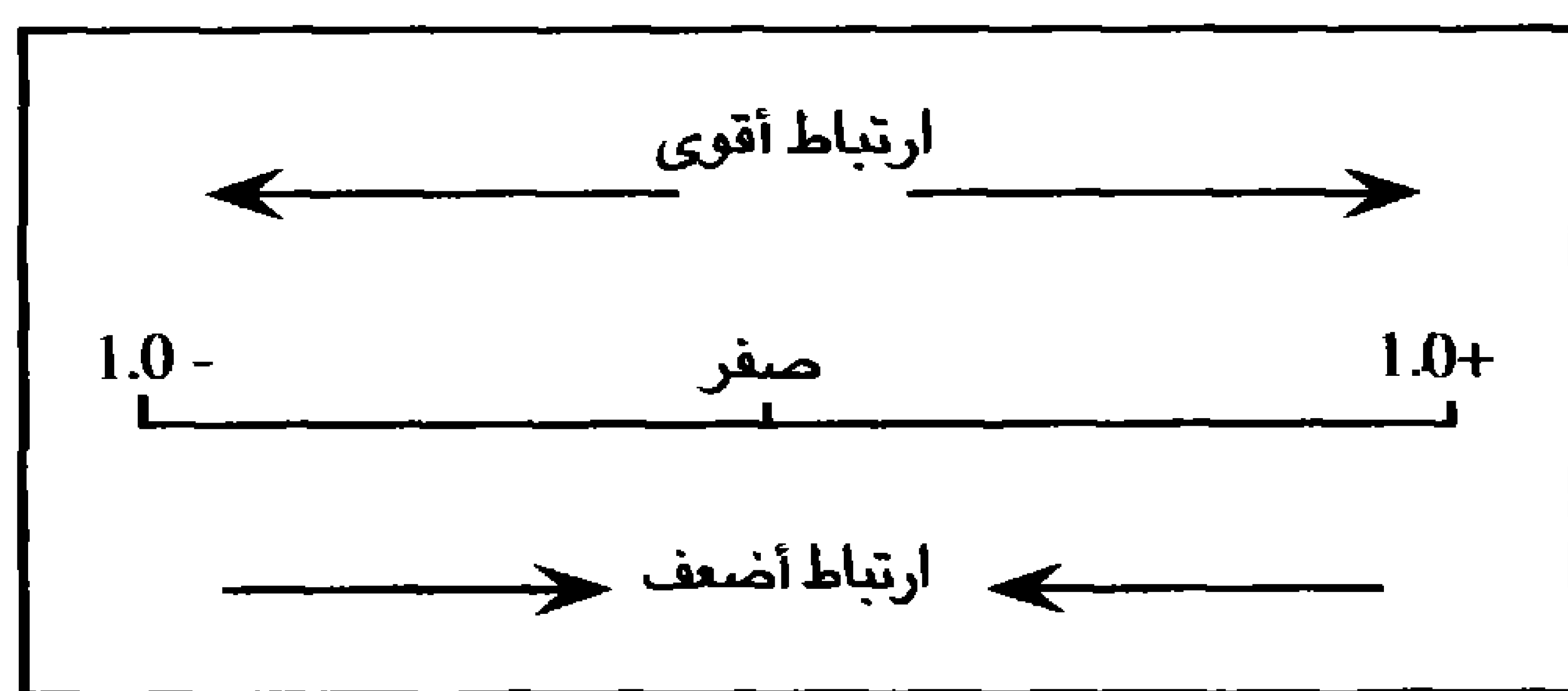
- نسبة 95% تقريباً من الأفراد تقع بين ±2 انحراف معياري.
- نسبة 99% تقريباً من الأفراد تقع بين ±3 انحراف معياري.

والتوزيع الاعتدالي له تطبيقات متعددة في القياس النفسي والتربوي، وبخاصة في تفسير درجات الاختبارات كما سيتضح في الفصل الرابع. غير انه ينبغي ملاحظة ان توزيعات درجات السمات والخصائص النفسية تتأثر تأثراً كبيراً بخصائص الاختبارات المستخدمة. لذلك فإنه ربما تختلف هذه التوزيعات عن التوزيع الاعتدالي او تقترب منه. فإذا افترضنا ان السمة تتوزع اعتدالياً، فإنه يمكن اجراء تحويل معين للدرجات المستمدة من الاختبارات، بحيث يصبح توزيع هذه الدرجات اعتدالياً.

مقاييس العلاقة:

نحتاج في القياس النفسي والتربوي في كثير من الأحيان إلى وصف العلاقة بين اختبارين. فالدرجات ربما تستمد من صيغتين مختلفتين للاختبار نفسه طبقتا على مجموعة واحدة من الأفراد، او ربما تستمد من اختبارين مختلفين يطبقان في وقت معين، او يطبق احدهما في احد الاوقات والآخر في وقت لاحق. وفي مثل هذه الحالات نود تحديد مقادر العلاقة بين مجموعتي الدرجات للتحقق من ثبات الدرجات، وصدق الاختبارات، كما سيتضح في الفصل الرابع.

ويمكن اجراء ذلك باستخدام معامل الارتباط Coefficient of Correlation الذي ينسب إلى كارل بيرسون Pearson ويرمز له بالرمز (ر). وتتراوح قيم هذا المعامل بين (+1.5، -1.0) كما هو موضح بالشكل (2 - 6) التالي:



: يوضح مدى قيم معامل الارتباط

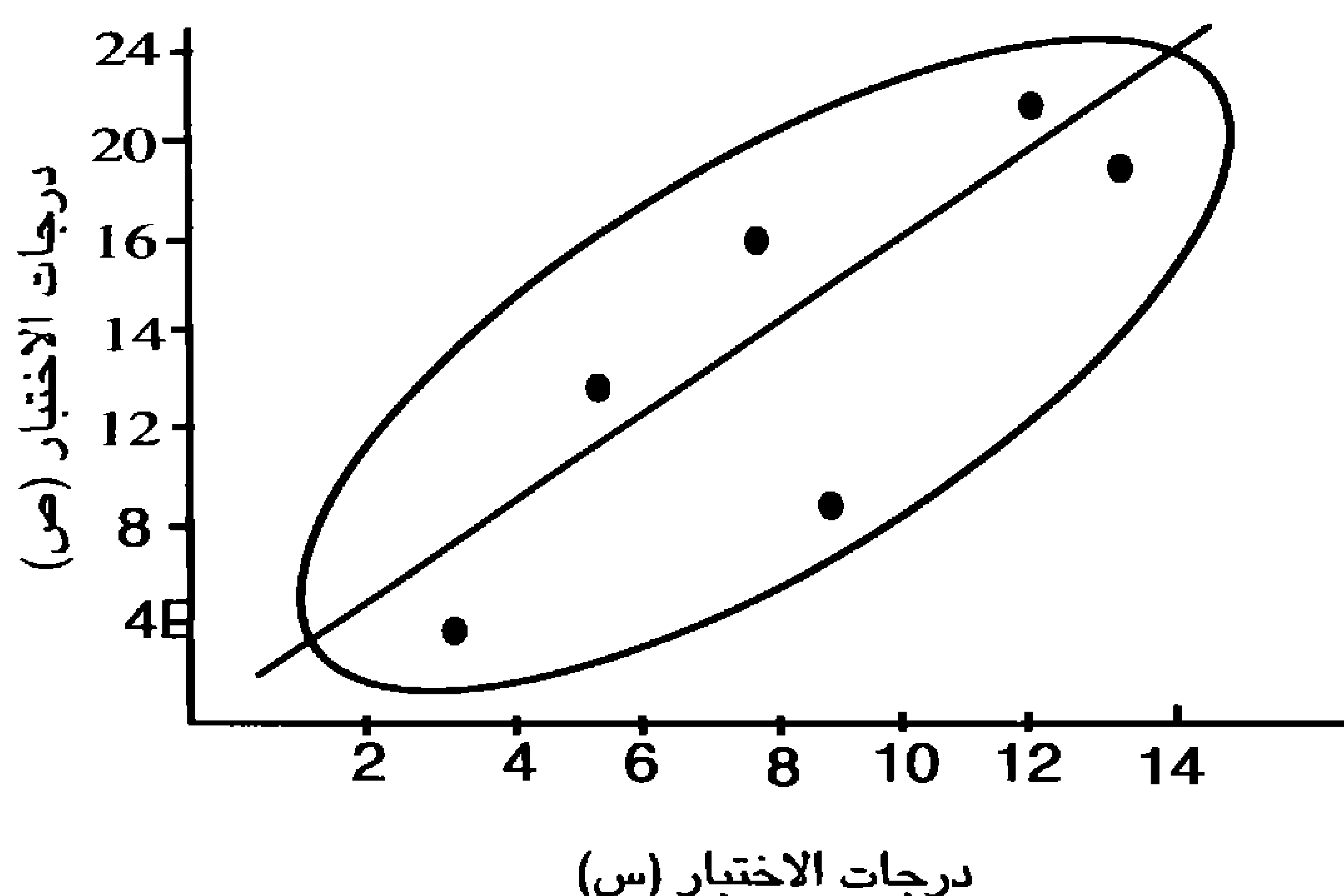
ويتضح من شكل (2 - 6) ان السهمين اعلى الخط الأفقي يشيران إلى ان الارتباط يصبح اقوى كلما ابتعدت قيمة المعامل عن الصفر نحو القيمة (+ 1) أو (-1) والسهمان اسفل الخط الافقي يشيران إلى ان الارتباط يصبح اضعف كلما اقتربت قيمة المعامل من الصفر الذي يعني عدم وجود علاقة بين درجات الاختبارين. وكلما ارتفعت قيمة معامل الارتباط سواء بالموجب أو السالب دل ذلك على ان الارتباط اقوى. ومقدار الارتباط يشير إلى درجة او قوة العلاقة، بينما تبين الاشارة الموجبة أو السالبة إلى اتجاه العلاقة. وعادة تكون قيمة معامل الإرتباط بين درجات الاختبارات والمقاييس التربوية والنفسية التي تستخدم في دراسة الفروق الفردية أقل من الواحد الصحيح. ولإلقاء الضوء على العلاقة بين درجات اختبارين مثلاً نرسم شكلاً انتشارياً لمجموعة الدرجات، وذلك بتحديد نقط تمثل الأزواج المتناظرة من الدرجات في كل من الاختبارين. وهذا يساعد في توضيح شكل العلاقة، اي خطية او منحنية، وأخذ فكرة مبدئية عن مدى قوة العلاقة واتجاهها.

ولتوضيح ذلك نفترض أن لدينا مجموعتين من الدرجات في اختبارين (س) ، (ص) كالتالي:

1 ، 3 ، 5 ، 7 ، 9 ، 11 ، 13

7 ، 4 ، 13 ، 16 ، 10 ، 22 ، 19

والشكل الانتشاري لهاتين المجموعتين يكون كالتالي:



تمثيل مجموعتي درجات اختبارين بشكل انتشاري:



ويتضح من شكل (2 - 7) ان الدرجات تميل إلى التراكم إلى حد ما حول خط مستقيم، وتتخذ الدرجات اتجاهاً إلى اعلى، مما يدل على ان العلاقة موجبة، وقوية بدرجة ما. وإذا وقعت جميع الدرجات على الخط المستقيم فإن معامل الارتباط يكون تاماً.

حساب قيمة معامل الارتباط.

لحساب مقدار العلاقة الخطية بين مجموعتين من درجات اختبارين طبقاً على عينة من الأفراد، يمكن استخدام الصيغة العامة لمعامل ارتباط بيرسون، وهي كالتالي:

ويتطلب تطبيق هذه الصيغة استخدام آلة حاسبة لايجاد قيم كل من: مجـ س ص، مجـ س، مجـ ص، مجـ $س^2$ ، مجـ $ص^2$.

فإذا استخدمنا هذه الصيغة لايجاد قيمة معامل الارتباط بين مجموعتي الدرجات السابقتين، فإننا نكون جدولاً كالتالي:

س	ص	$س^2$	$ص^2$	س ص
1	7	1	49	7
3	4	9	16	12
5	13	25	169	65
7	16	49	256	112
9	10	81	100	90
11	22	121	484	242
13	19	169	361	247
المجموع 49	91	455	1435	775

$$ر = \frac{ن\ مجـ\ س\ ص - مجـ\ س \times مجـ\ ص}{\sqrt{[ن\ مجـ\ س^2 - (مجـ\ س)^2][ن\ مجـ\ ص^2 - (مجـ\ ص)^2]}}$$

$$= \frac{(91 \times 49) - (775 \times 7)}{\sqrt{[7 \times 455 - (49)^2]\ [7 \times 1435 - (91)^2]}}$$

$$= \frac{5425 - 4459}{\sqrt{784 \times 1764}}$$

$$= \frac{966}{28 \times 42} = \frac{23}{28} = +0.82\ .$$

ويلاحظ ان قيمة (ر) تنحصر بين الصفر والواحد الصحيح، ولكنها اقرب إلى الواحد الصحيح، مما يدل على أن العلاقة موجبة وقوية إلى حد ما، وهذا يتضح من الشكل الانتشاري الذي تراكمت فيه درجات الاختبارين حول خط مستقيم. كما انها قيمة موجبة لأن الدرجات اتخذت اتجاهاً إلى اعلى في الشكل.

ولتفسير قيمة معامل ارتباط بيرسون ينبغي ايجاد نسبة التباين في درجات الاختبار (س) التي يمكن ان ترجع إلى تباين درجات الاختبار (ص)، اي التباين المشترك بين درجات الاختبارين، الذي يساوي مربع قيمة معامل الارتباط ($ر^2$) .

ففي المثال السابق $ر^2 = (0.82)^2 = 0.67$ مما يعني ان نسبة 67% من التباين في درجات الاختبار (س) تعزى إلى تباين درجات الاختبار (ص) أما التباين الباقي الذي يساوي $1.00 - 0.67 = 0.33$ فإنه يعزى إلى عوامل اخرى ربما ترجع إلى الأفراد المختبرين، مثل: تذبذب الدافعية، او الذكاء، او إلى عوامل خارجية تتعلق بالبيئة او ظروف تطبيق الاختبارين، او غير ذلك من العوامل غير المحددة. لذلك يطلق على ($ر^2$) معامل التحديد Coefficient of Determination، ويطلق على ($1 - ر^2$) معامل الاغتراب او عدم التحديد Coefficient of Alienation.

ومن الجدير بالذكر انه على الرغم من تعدد استخدامات معامل الارتباط في القياس النفسي والتربوي، الا انه يساء في كثير من الاحيان استخدام وتفسير هذا المعامل. ويرجع



ذلك لأسباب متعددة يصعب مناقشتها في هذا الفصل المحدود. ويمكن ان يرجع القارئ إلى كتاب تحليل البيانات في البحوث النفسية والتربوية والاجتماعية للمؤلف للاطلاع على تفاصيل ذلك.

الاعتماد على ($ر^2$) في تفسير قيمة معامل الارتباط (ر). فلا يجب مثلاً اعتبار ان القيمة 0.50 تعني ارتباطاً متوسطاً، وانما تعني ان التباين المشترك بين المتغيرين 0.25، مما يدل على ان قيمة معامل الارتباط 0.50 منخفضة بينما، قيمة (ر) التي تساوي 0.71 تقريباً تجعل التباين المشترك 0.50، مما يعني ان ر = 0.71 قيمة متوسطة.

تتأثر قيمة معامل الارتباط بين متغيرين بعوامل كثيرة مثل: تضييق مدى الدرجات او تشتتها، وأخطاء القياس، وشكل توزيع درجات كل من المتغيرين، ومدى تجانس العينة، وانحناء العلاقة بدلاً من ان تكون خطية (مستقيمة)، وغير ذلك.

استخدام معامل ارتباط بيرسون يتطلب أن يكون مستوى قياس كل من المتغيرين المراد ايجاد العلاقة بينهما فترياً او نسبياً. وهناك معاملات ارتباط اخرى تناسب المتغيرات من المستويين الاسمي والرتبي يمكن الرجوع اليها في الكتاب المشار إليه أعلاه.

الخلط في تفسير قيمة معامل الارتباط بين الاقتران او العلاقة، والعلِّية Causality؛ فالقيمة المرتفعة لمعامل الارتباط لا تعني العلِّية او السببية، او علاقة أثر ونتيجة. فتباين درجات الاختبار الأول ليست سبباً في تباين درجات الاختبار الثاني مهما بلغت قيمة معامل الارتباط بين مجموعتي الدرجات، إذ ربما تكون هذه العلاقة راجعة إلى ان كلاً من الاختبارين يرتبطان بمتغير ثالث إذا عزلنا اثره قد تنعدم هذه العلاقة؛ فالارتباط يعني انه يمكن التنبؤ بأحد المتغيرين باستخدام الآخر بدرجة معينة من الثقة، أو أن هناك اطراداً بين المتغيرين.

يمكن استخدام معامل ارتباط بيرسون ليس فقط في ايجاد مقدار العلاقة الخطية بين درجات اختبارين، وانما يفيد ايضاً في التوصل إلى معادلة افضل خط مستقيم كما في الشكل الانتشاري (2 - 7) لأزواج الدرجات، وهو الخط الذي تقترب منه جميع النقط الممثلة

لهذه الدرجات، ويسمى خط الانحدار Regression line لأن النقط تكون متراكمة حوله، وتميل إلى ان تنحدر واقعة عليه.

ومعادلة خط الانحدار للمتغير (ص) على المتغير (س) كالتالي:

حيث ترمز (ص م) إلى قيمة (ص) المتنبأ بها.

، (س̅) ، (ص̅) إلى متوسط كل من المتغيرين (س) ، (ص) على الترتيب.

، (ع س)، (ع ص) إلى الانحراف المعياري لكل من المتغيرين (س) ، (ص) على الترتيب.

، (ر) إلى معامل ارتباط بيرسون.

وينتج عن تطبيق هذه المعادلة ايجاد قيمة (ص) المراد التنبأ بها، حيث يتم التعويض عن جميع الرموز الأخرى في المعادلة بقيمها المعلومة المناسبة.

وتُستخدم هذه المعادلة بكثرة في أغراض الانتقاء، وتقدير الصدق التنبؤي للاختبارات والمقاييس التربوية والنفسية كما سيتضح في الفصل الرابع.

ويتضح من ذلك ان معامل الارتباط هو المقياس الاحصائي الذي يستخدم في التنبؤ، وتفسير العلاقات بين الظواهر التربوية والنفسية، مما يجعله مقياساً ضرورياً لدراسة الفروق الفردية، وكشف العلاقة بين متغيراتها.

الفصل الثالث

الدرجات المحوَّلة، ومعايير الاختبارات

- مقدمة
- أنواع معايير الاختبارات
- الدرجات التائية
- المئينيات، والإعشاريات
- التساعيات المعيارية
- نسبة الذكاء
- نسبة الذكاء الانحرافية
- معايير الفرق الدراسية
- الصفحات النفسية (البروفيلات)

مقدمة:

أوضـحنا فيما سبق أن الاختبارات التربوية والنفسية هي ادوات قياس مقننة لعينة ممثلة من السلوك بطريقة منتظمة وفقـاً لقواعد مـحددة، وتسـتـخدم نتائج القياس في المقـارنة بين الافـراد وفـقـاً لمعيـار أو مـيـزان مـعـرف تعـريفـاً واضـحـاً، مما يسـاعـد في تفـسـيـر درجـات الاختبارات. فالدرجات التي يحصل عليها الافراد في اختبار معين تعبر عن عدد المفردات أو الأسئلة التي اجاب عليها كل منهم اجابة صـحيحة، وتسـمى الدرجات الخام Raw Scores، وهذه الدرجـات يصـعب تفـسيرهـا مـا لم يتـفق على اطار مـرجـعي Refrencing Scheme يتم على اساسه تفسير الدرجات. فالدرجة (115) التي حصـــل عليـــها فرد في اختبـــار للذكاء لا تعني شيئاً مـا لم تُنسب إلى معيار أو محك مـعين. إذ ربما يتبادر إلـى الأذهان أن هذا الفرد اجاب عن (115) سؤالاً اجابة صحيحة، أو ربما نقول انه حصل على (137) نقطة من المجموع الكلي من النقط وهو (200) مثلاً. ولكن هذه الدرجة لا تنقل لنا صورة واضحة عن مركز الفرد بالنسبة لأقرانه في الذكاء. فمثلاً لو أن الدرجة (115) التي حصل عليها الفرد تفوق 65% من درجات اقرانه الذين يتفقون مـعـه في العمـر، والمسـتوى التـعليمي، والمسـتوى الاجـتمـاعي والاقـتصـادي وغيـر ذلك، فـإننا نكون قـد فسـرنا هـذه الدرجـة في اطار الجـمـاعة المرجـعيـة Refrence Group التي ينتمي اليها هذا الفرد.

وهذا يدل على أن تفسير درجات الاختبارات والمقاييس المقننة لا يكون تفسيراً مطلقاً كما هو الحـال في القيـاس الفـيزيائي، وانما يكون بارجـاع الدرجـة إلى مـعـيـار الجـمـاعة Group Norm، أي انه تفـسـيـر نسـبي. وتوجـد مـواقف قليلة في القـيـاس النفـسي والتـربوي يمكن تفسيرها استناداً إلى محكات أو معايير مطلقة Absolute Standards؛ فإذا قلنا أن فرداً ما استطاع أن يجيب عن 75% من أسئلة اختبار تقيس مهارة حسابية معينة، وتم تحديد النطاق الشامل للأسئلة التي يمكن أن ينتقي منها اسئلة الاختبار بطريقة عشوائية بحيث تمثل هذا النطاق تمثيلاً جيداً، فـإنه يمكن في هـذه الحـالة مـقارنة درجـة الفرد في الاخـتبـار بمحك أو

مستوى أداء مطلق يمثله النطاق الشامل للمهارة المقاسة. ولكننا لا نستطيع تحديد ما اذا كانت النسبة 75% جيدة أو متوسطة أو ضعيفة بالنسبة لمن هم في عمر هذا الفرد، وفي مستوى تعليمه وتدريبه على هذه المهارة. فهذه المقارنة تتطلب معرفة أداء مجتمع هؤلاء الافراد في المهارة الحسابية، لكي نحدد مركز الفرد في هذه المهارة بالنسبة لهذه الجماعة المعيارية.

لذلك فإن المعايير Norms تعد الأساس الذي يُستند اليه في تفسير درجات الاختبارات والمقاييس المقننة، حيث يمكن في ضوئها مقارنة أداء الفرد في الاختبار بأداء اقرانه، أي الجماعة المرجعية المعيارية Standard Group. ونظراً لوجود جماعات مرجعية مختلفة يمكن استخدامها في الحصول على معايير الاختبارات، فإن تحديد الجماعة المرجعية المناسبة يعد امراً ضرورياً. فإذا كان الاختبار قد صُمِّم لقياس الاستعداد الدراسي للطلاب الذين يودون الالتحاق بكلية معينة، فإن الجماعة المرجعية يجب أن تشتمل على طلاب الثانوية العامة الذين سوف يلتحقون بهذه الكلية، وإذا كان الهدف من الاختبار قياس سمات شخصية معينة لدى المراهقين من البنين مثلاً، فان الجماعة المرجعية يجب أن تشتمل على عينة مستعرضة من المراهقين البنين، وهكذا. أما اذا كان للاختبار أكثر من غرض، أو صُمِّم لكي يستخدم مع جماعات متعددة، فإن هذا سوف يتطلب اكثر من جماعة مرجعية واحدة. فمدى ملاءمة جماعة مرجعية معينة يعتمد على التشابه بين المجتمع الذي تتم معاينته والمجموعة التي يود مستخدم الاختبار مقارنة درجته بها.

أنواع معايير الاختبارات:

يتضح مما سبق أن الدرجة الخام للفرد في الاختبار لا تفيد شيئاً في تفسير هذه الدرجة. لذلك ينبغي اجراء نوع من التحويل على هذه الدرجة لكي تصبح قابلة للمقارنة، وتحدد مكانة الفرد أو مركزه بالنسبة لأقرانه في الجماعة المرجعية. ومن هنا تأتي اهمية معايير الاختبارات Tests Norms. فدرجات اختبارات الذكاء، والاستعدادات، والشخصية، والاختبارات التحصيلية المقننة، وغيرها تفسَّر درجاتها عادة استناداً إلى جداول المعايير Norms Tables التي تكون مدونة في أدلة الاختبارات. ويمكن باستخدام هذه الجداول معرفة الدرجة المحوَّلة المناظرة لدرجة خام معينة. وهذه الدرجات المحوَّلة أو المعايير يتم اعدادها إمبريقياً، وذلك بتحديد أداء عينة من الأفراد ممثلة لمجتمع معين في الاختبار، أي الجماعة المرجعية. فإذا اشتملت هذه الجماعة على عينة قومية مستعرضة مستمدة من جميع مناطق بلد معين بحيث

تمثل هذه المناطق، فإن المعايير عندئذ تسمى المعايير القومية National Norms، وهذه المعايير شائعة الاستخدام في كثير من الاختبارات المقننة لأنها تتميز بقابلية التعميم.

أما اذا اشتملت الجماعة المرجعية على عينة من منطقة أو محافظة معينة، فإن المعايير المستمدة منها تسمى عندئذ المعايير المحلية Local Norms. وتفسير درجات الاختبارات استناداً إلى هذه المعايير يكون محدوداً بهذه العينة المحلية، ولكنها تكون اكثر فائدة من المعايير القومية اذا اردنا مقارنة أداء الفرد بأداء افراد آخرين مماثلين له في الخبرة. كما أن بعض الاختبارات يكون لها معايير منفصلة للذكور والإناث، أو الصفوف المدرسية، أو العمر. كما أن بعض مقاييس الميول تستند إلى معايير منفصلة للمجموعات المهنية المختلفة، وبعض اختبارات القبول بالدراسات العليا بالجامعات تستند إلى معايير منفصلة للتخصصات الاكاديمية المختلفة.

فالمعايير اذن هي مجموعة من الدرجات التي تشتق بطرق إحصائية معينة من الدرجات الخام، بحيث تأخذ في الاعتبار توزيع الدرجات الخام التي نحصل عليها نتيجة تطبيق الاختبار على عينة عشوائية ممثلة لمجتمع معين من الأفراد، وأحياناً تسمى عينة التقنين Standardization Sample. وتصمم هذه المعايير لتحقيق هدفين: أحدهما تحديد مكانة أو مركز الفرد بالنسبة لعينة التقنين، وبذلك يمكن تقييم ادائه بالنسبة لأقرانه، والآخر تقديم قياسات متقارنة، أي تسمح بالمقارنة المباشرة بين درجات الفرد في اختبارات مختلفة؛ فمثلاً إذا حصل الفرد على الدرجة 25 في اختبار في الاستدلال الحسابي، والدرجة (14) في اختبار فهم معاني الكلمات، فإنه يصعب باستخدام هاتين الدرجتين معرفة المركز النسبي للفرد في كل من الاختبارين، أي لا نستطيع معرفة ما إذا كان أفضل في الاستدلال الحسابي أم في فهم معاني الكلمات. ويرجع ذلك إلى أن مستوى صعوبة مفردات الاختبار تؤثر في إمكانية المقارنة بين الدرجتين الخام. أما الدرجات المحولة فإنه يُعبَّر عنها بنفس الوحدات، وتستند إلى نفس عينة تقنين كل من الاختبارين، مما يمكننا من مقارنة أداء الفرد في اختبارات مختلفة.

وتوجد طرق متعددة لإجراء تحويل للدرجات الخام تؤدي إلى أنواع مختلفة من المعايير التي تُستخدم في تفسير درجات الاختبارات التربوية والنفسية أهمها ما يلي:

الدرجات المعيارية Standard Scores :

تعد الدرجات المعيارية من التحويلات الخطية للدرجات الخام التي تعتمد على المتوسط

والانحراف المعياري لهذه الدرجات، وتؤدي إلى توزيع للدرجات المحوَّلة مطابق لتوزيع الدرجات الخام. وهذا يعني أنه إذا مثَّلنا توزيع الدرجات الخام الأصلية على المحور الأفقي، والدرجات المعيارية على المحور الرأسي في شكل بياني، فإن النقط المناظرة لنوعي الدرجات سوف تقع على خط مستقيم.

والدرجات المعيارية هي عدد الانحرافات المعيارية لدرجات الاختبار عن متوسط هذه الدرجات لمجموعة معينة من الأفراد. فإذا حصل فرد على الدرجة (120) في اختبار متوسط درجاته (100)، وانحرافه المعياري (10) فإن الدرجة المعيارية المناظرة لهذه الدرجة تساوي $\frac{120-100}{10}$ = +2. فالدرجة الخام (120) تنحرف عن المتوسط (100) بقدر (20) درجة، ولكن (20) درجة تعني 20 ÷ 10 = + 2 انحراف معياري. أي أن درجة هذا الفرد تبعد بقدر انحرافين معياريين موجبين عن متوسط الجماعة المرجعية أو عينة التقنين، أو أن درجة الفرد تفوق متوسط درجات أقرانه بقدر انحرافين معيارين قيمة كل منهما (10) درجات.

وعندما نحوِّل جميع درجات اختبار إلى درجات معيارية، فإن التوزيع الناتج يكون متوسطه صفراً، وانحرافه المعياري مساوياً الواحد الصحيح. وذلك لأن الدرجة المعيارية صفر تدل على أن الدرجة الخام لا تنحرف عن المتوسط، أي تساوي المتوسط، والدرجة المعيارية (+1) تدل على أن الدرجة الخام تزيد عن المتوسط بقدر انحراف معياري واحد، والدرجة المعيارية (-1) تدل على أن الدرجة الخام تقل عن المتوسط بقدر وحدة انحراف معياري.

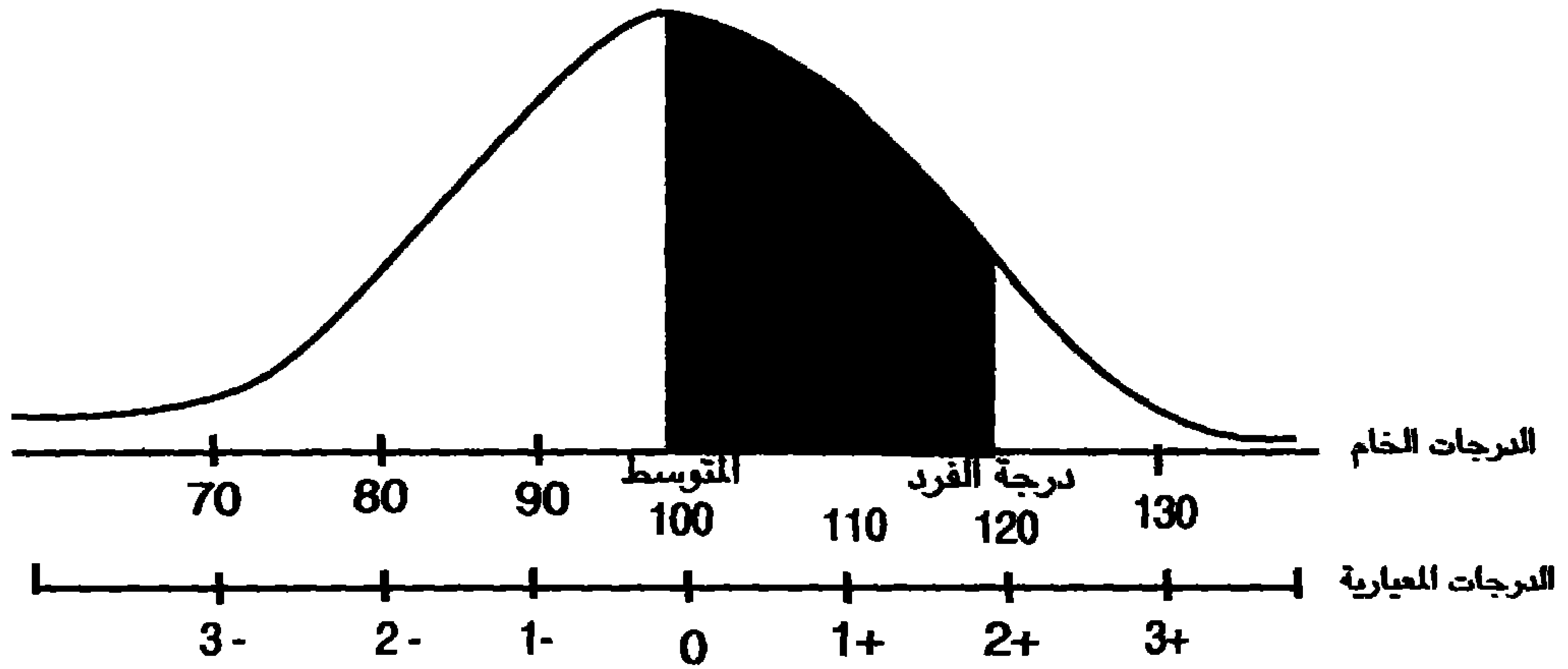
ومدى الدرجات المعيارية لاختبار معين ينحصر عادة بين (+3) ، (-3) انحراف معياري.

والصيغة المستخدمة في حساب الدرجات المعيارية كالتالي:

$$\text{الدرجة المعيارية} = \frac{\text{الدرجة الخام} - \text{المتوسط}}{\text{الانحراف المعياري}}$$

أي أن : $د = \frac{س - \bar{س}}{ع}$ ...... (3 - 1)

ويمكن تمثيل الدرجة الخام (120) المذكورة في المثال السابق على المنحنى الاعتدالي كما هو مبين بشكل (3 - 1) التالي:



شكل (3 - 1): يوضح توزيع الدرجات المعيارية على المنحنى الاعتدالي

ومن هذا يتضح أن الدرجات المعيارية تدل على بُعد الدرجة الخام عن متوسط توزيع الدرجات، وأن الدرجات التي تفوق المتوسط تكون اشارتها موجبة، والتي تقل عن المتوسط تكون إشارتها سالبة. ونظراً لأن الدرجات المعيارية يُعبر عنها على ميزان فترى، فإنه يمكن اجراء العمليات الحسابية عليها، وكذلك فإنه نظراً لأن تحويل الدرجات الخام إلى درجات معيارية يُعد تحويلاً خطياً Linear Transformation، فإن شكل توزيع الدرجات المعيارية سوف يكون مماثلاً لتوزيع الدرجات الخام كما سبق أن أوضحنا؛ فإذا كان توزيع الدرجات الخام اعتدالياً، فإن توزيع الدرجات المعيارية سيكون اعتدالياً كذلك، وإذا اقترب توزيع الدرجات الخام من التوزيع الاعتدالي، فإن مدى الدرجات المعيارية سوف يتراوح تقريباً بين (3-) ، (3+) .

الدرجات التائية (ت) T scores :

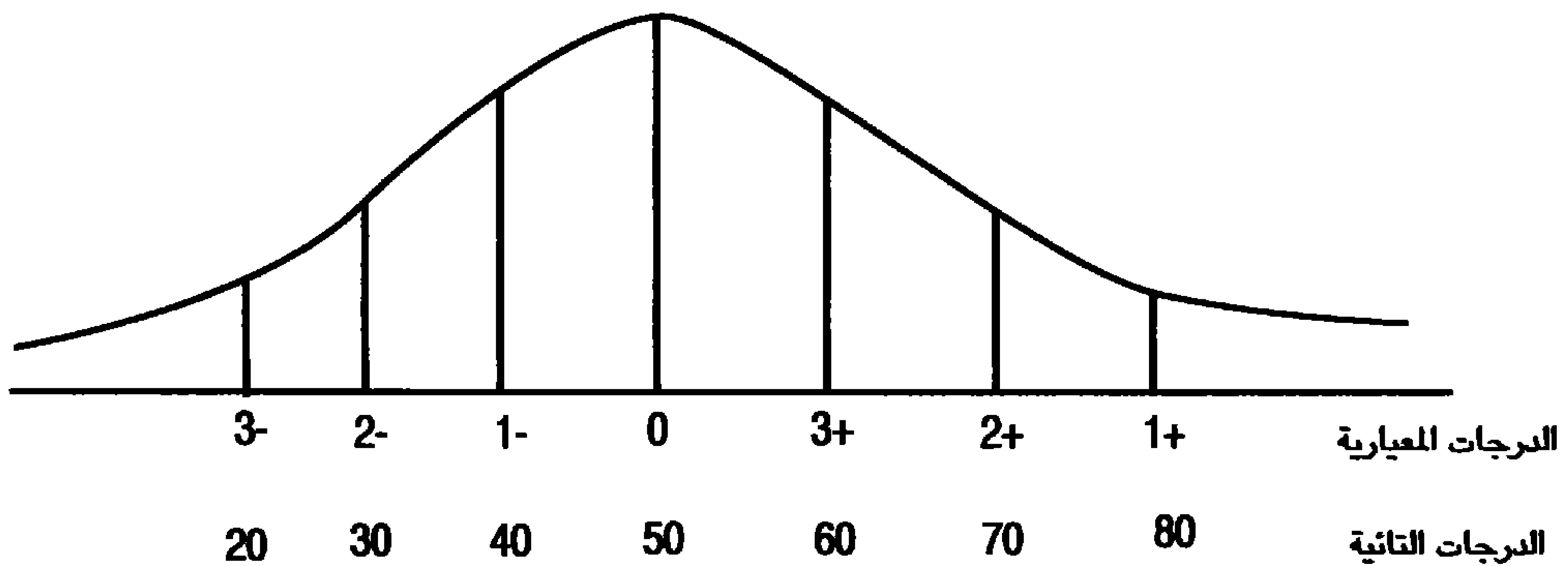
على الرغم من ميزات الدرجات المعيارية، إلا انه يُعاب عليها أنها تشتمل على كسور وإشارات سالبة. وللتغلب على ذلك يجرى عادة تحويل ميزان الدرجات المعيارية إلى ميزان آخر مناسب. وهذا يماثل تحويل درجات الحرارة المئوية إلى درجات فهرنهيتية أو العكس. وقد اقترح مكول (McCall, 1922) أن نضرب الدرجة المعيارية في (10) للتخلص من الكسور، وإضافة (+ 50) إلى الناتج للتخلص من الإشارات الجبرية السالبة. وبذلك نحصل على درجة

محوّلة جديدة تسمي الدرجة التائية (ت). وقد اختار مكول McCall الحرف (t) أو (ت) تيمُّناً بكل من رائدي علم النفس ثورنديك (Thorndike) وتيرمان (Terman) حيث إنه الحرف الأول من اسميهما.

أي أن: ت = 10 د + 50 ...... (3 - 2)

فالدرجة التائية المناظرة للدرجة المعيارية (-1.3) تساوي 10 (-1.3) + 50 = 37، والدرجة التائية المناظرة للدرجة المعيارية (صفر) تساوي 10 (صفر) + 50 = 50. وهكذا في بقية الدرجات.

وعندما تُستخدم الدرجات المعيارية التي توزيعها اعتدالي، فإننا نحصل على توزيع الدرجة التائية التي تتخذ شكل التوزيع الاعتدالي أيضاً، ويكون متوسطها 50 وانحرافها المعياري 10، كما هو موضح بالشكل (3 - 2) التالي:



شكل (3 - 2): يوضح ميزان الدرجات المعيارية والتائية

ومن الجدير بالذكر أنه يمكن إجراء تحويلات خطية أخرى على الدرجات المعيارية للحصول على ميزان جديد للدرجات يناسب أغراضاً معينة. فمثلاً اختبار القبول بالدراسات العليا بالجامعات الامريكية Graduate Record Examination (GRE)، وكذلك اختبار القبول بالجامعات College Entrance Examination Board (CEEB) متوسط درجات كل منهما 500 وانحرافه المعياري 100. أي أن:

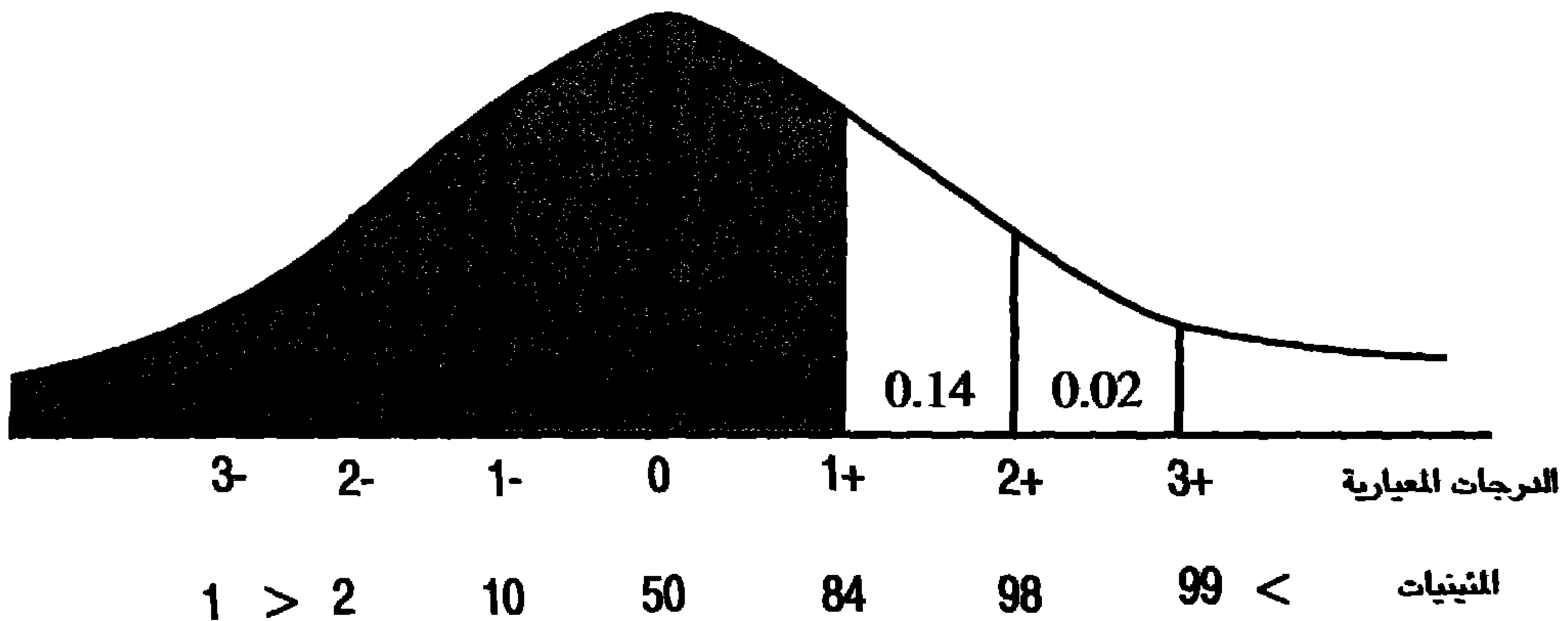
الدرجة المحولة = 100 د + 500 .... (3 - 3)

المئينيات Percentiles، والإعشاريات Deciles :

تعتمد كثير من الاختبارات المقننة في تفسير درجاتها على المئينيات، أو الرتب المئينية Percentile Ranks. فالمئيني هو نقطة على توزيع الدرجات تقابلها أو تقل عنها نسبة معينة من الأفراد. فالمئيني 50 مثلاً هو النقطة التي تقسم التوزيع إلى نصفين متساويين، وهذه النقطة تناظر وسيط الدرجات. وإذا كانت الدرجة الخام التي حصل عليها فرد في الاختبار تناظر المئيني (72)، فإن هذا يعني أنه يفوق 72% من أفراد الجماعة المرجعية المعيارية أو عينة التقنين، ويقل عن 28% من هؤلاء الأفراد. لذلك فإن المئينيات تحدد بطريقة مباشرة المركز النسبي للفرد في مجموعته.

ويمكن إيجاد المئينيات المناظرة للدرجات الخام في توزيع معين بطريقة بيانية بإيجاد التكرار المتجمع الصاعد المئوي لكل فئة من فئات التوزيع، ورسم المنحنى التكراري المتجمع المئوي كما سبق أن أوضحنا في الفصل الثاني. ورسم خط رأسي من الدرجة التي تقع على المحور الأفقي حتى يلاقي المنحنى في نقطة أخرى نسقط منها عموداً على المحور الرأسي، ونقطة الالتقاء بهذا المحور تناظر المئيني المطلوب.

ويمكن توضيح العلاقة بين الدرجات المعيارية والمئينيات على المنحنى الاعتدالي المعياري بالشكل (3 - 3) التالي:



شكل (3 - 3) يوضح العلاقة بين الدرجات المعيارية والمئينيات

ويتضح من شكل (3 - 3) أن الدرجة المعيارية (+1) تفوق المتوسط بقدر انحراف معياري واحد، وتناظر المئيني 84. والدرجة المعيارية (-1) تقل عن المتوسط بقدر انحراف معياري واحد، وتناظر المئيني 14. وقد حصلنا على المئيني 84 بإضافة 50% من الحالات التي تقل عن المتوسط إلى 34% من الحالات التي تنحصر بين الدرجتين المعياريتين صفر، + 1.

وعلى الرغم من أن الدرجات المعيارية في التوزيع الاعتدالي يكون مستوى قياسها فترياً، وذلك لأن الفرق بين كل درجتين معياريتين متساوٍ، إلا أن هذا لا ينطبق على المئينيات؛ فالفرق بين المئيني 50 والمئيني 84 يناظر فرقاً قدره انحراف معياري واحد (د = صفر، د = +1) كما هو مبين بالشكل، ويحد نسبة من المساحة تحت المنحنى الاعتدالي قدرها 34% .

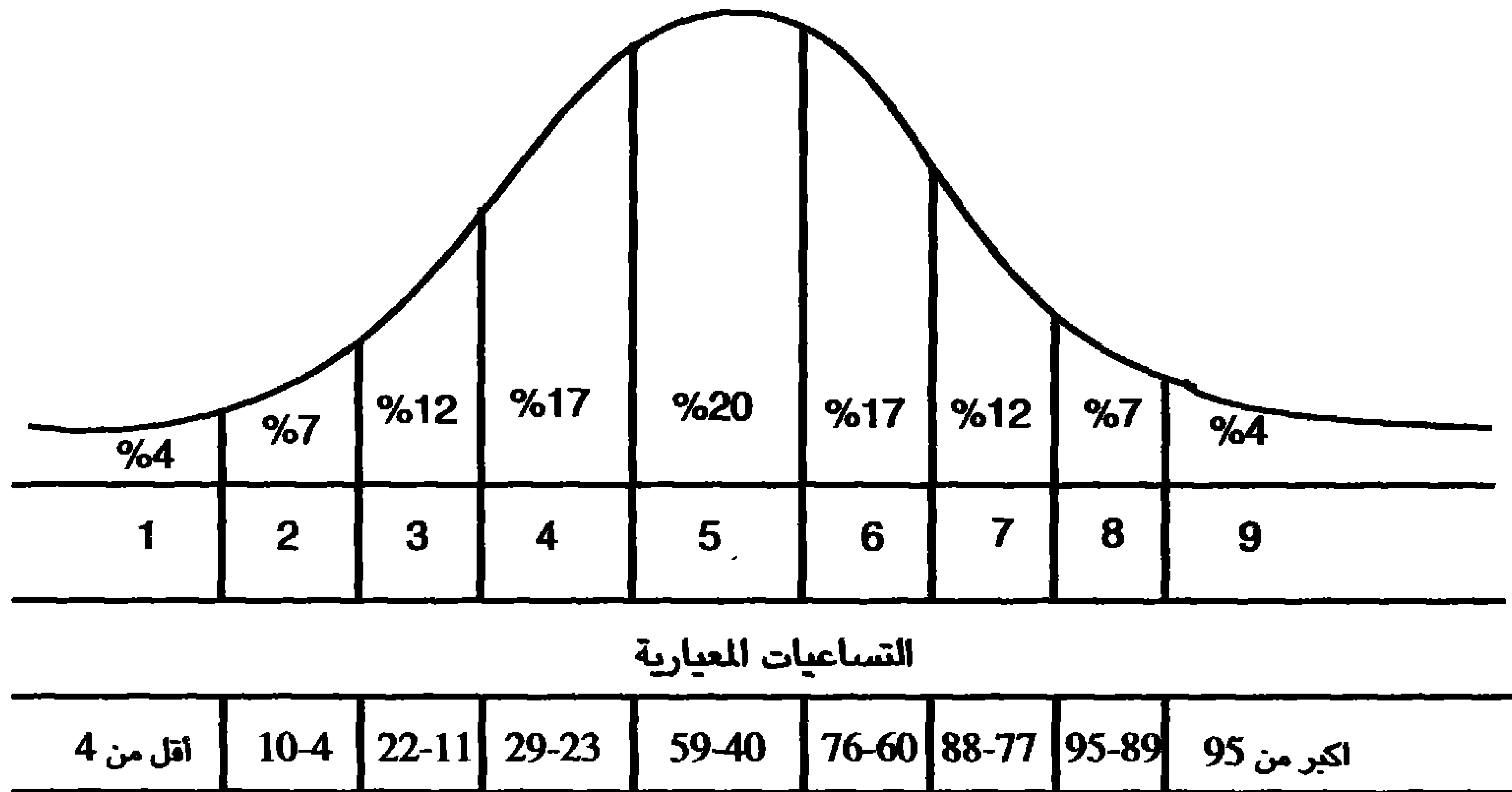
بينما نجد أن الفرق بين المئيني 84 والمئيني 98 يناظر أيضاً فرقا قدره انحراف معياري واحد (د = + 1 ،د = + 2) ولكن يحد نسبة من المساحة تحت المنحنى الاعتدالي قدرها 14% فقط.

فالمئينيات مستوى قياسها يكون رتبياً، ولذلك لا نستطيع إجراء العمليات الحسابية الأربع عليها. وهذا يعد أمراً أساسياً عند تفسير درجات الاختبارات، حيث إن الفرق بين مئيني وآخر بالقرب من وسط التوزيع يكون اقل بكثير من الفرق نفسه بالقرب من طرفي التوزيع. فالطالب الذي تحسن أداؤه بحيث تغير المئيني المناظر لدرجته في الاختبار من المئيني (1) إلى المئيني (5) مثلاً يكافى تحسن طالب آخر من المئيني (50) إلى المئيني (70). ولعل هذا يؤدي إلى صعوبة المقارنة بين المئينيات لاختلاف الميزان من موقع إلى آخر على متصل الدرجات. ويمكن تقسيم المساحة تحت المنحنى الاعتدالي المعياري إلى (10) أقسام بدلاً من (100)، وبذلك يمكن التعبير عن المئينيات بالإعشاريات Deciles. فالمئيني (70) مثلاً يناظر الإعشاري (7)، وهكذا. ويمكن استخدام الإعشاريات إذا لم يكن هناك اهتمام بالفروق الدقيقة بين الأفراد.

التساعيات المعيارية Stanines :

يمكن إجراء تحويلات خطية متعددة على الدرجات المعيارية الاعتدالية بحيث يصبح متوسطها وانحرافها المعياري مناسباً للغرض من الاختبار. وأحد هذه التحويلات الشائعة التي استُخدمت بكثرة في الاختبارات التي صممتها هيئة القوات الجوية الامريكية أثناء الحرب العالمية الثانية، وفي أغراض تقدير درجات الطلاب في مؤسسات تربوية معينة، أطلق عليها

"التساعيات المعيارية Stanines" وهو اسم مكون من ضم كلمتي Standard و nine. وميزان التساعيات Stanine Scale هو درجة معيارية اعتدالية متوسطها (5) وانحرافها المعياري (2) تقريباً، استناداً إلى أن الدرجات تتراوح من (1) إلى (9). لذلك فإن التساعيات المعيارية يعبر عنها بأعداد ذات رقم واحد. ويمكن تحويل الدرجات الخام إلى تساعيات معيارية، وذلك بترتيب هذه الدرجات ترتيباً تصاعدياً أو تنازلياً، وتعيين التساعيات وفقاً للنسب المئوية من المساحات تحت المنحنى الاعتدالي. وتوجد تسع نقاط درجات أو تساعيات على ميزان القياس. ولكن ميزان التساعيات المعيارية لا يعد ميزان درجات معيارية حقيقي، وذلك لأن كلاً من التساعي الأول والتاسع مفتوح النهاية كما هو مبين بالشكل (3 - 4) التالي:



شكل (3 - 4) : يوضح التساعيات المعيارية، ومدى المئينيات ، ونسبة عدد الحالات

ويتضح من الشكل أن الدرجات التي تقع في حدود كل منطقتين يُعين لها أحد التساعيات، وكل منطقة اتساعها نصف انحراف معياري، فيما عدا كل من التساعي الأول والتاسع. وتحدد النسب المئوية من الحالات في المنحنى الاعتدالي داخل كل من منطقة، وكذلك مدى المئينيات المناظرة لكل منها؛ فجميع الأفراد الذين تقع درجاتهم في مدى معين يُعين لهم التساعي المعياري المناظر لهذا المدى. فمثلاً الأفراد الذين تقع درجاتهم الخام في المدى المئيني (60 - 76) يُعين لهم التساعي المعياري (6) وهكذا، حيث يُعين لنسبة 17% من افراد عينة التقنين التساعي المعياري (6) .



وتوصي بعض المؤسسات المعنية بنشر الاختبارات مرجعية المعيار Norm - Referenced Tests مثل: الاختبارات التحصيلية المقننة، واختبارات الاستعدادات، في كثير من الأحيان باستخدام التساعيات المعيارية في تفسير درجات هذه الاختبارات. وربما يرجع ذلك إلى أن التساعيات المعيارية تتعامل برقم واحد، كما انها تتعامل مع مدى المئينيات بدلاً من نقط المئينيات، وذلك لأن درجات الاختبارات يشوبها عادة قدراً من الخطأ، كما أنه نظراً لأن التساعيات المعيارية تعد درجات معيارية اعتدالية، فإنه يمكن إجراء معالجات رياضية عليها، فهي على العكس من المئينيات يمكن إيجاد متوسطها دون الحاجة إلى تحويلها إلى موازين أخرى.

غير أن بعض خبراء القياس يرى أن تفسير درجات الاختبارات باستخدام التساعيات المعيارية يعد أكثر صعوبة من تفسيرها باستخدام المئينيات، وبخاصة إذا كانت الدرجات تتميز بالثبات، حيث أن التساعيات المعيارية لا تتعامل مع كل درجة على حدة، وانما مع مجموعة من الدرجات في فئة معينة كما يتضح من شكل (3 - 4). لذلك يرى هؤلاء الخبراء انه يُفضل استخدام عبارات لفظية لوصف كل تساعي معياري. فمثلاً التساعي المعياري الأول يعني مستوى أدنى (4%) ، والتساعي المعياري الثاني يعني مستوى ضعيف (7%)، والتساعي المعياري الثالث يعني مستوى أقل من المتوسط (12%)، وهكذا. غير أن هذه العبارات اللفظية تنطبق فقط إذا كانت الدرجات تتوزع اعتداليا، وذلك لأنه إذا كان الاختبار غاية في السهولة أو الصعوبة، فإن توزيع الدرجات في هاتين الحالتين ربما يكون ملتوياً التواءً ملحوظاً، مما يتطلب الحيطة في تفسير التساعيات المعيارية، وبخاصة في طرفي التوزيع.

نسبة الذكاء The Intelligent Quotient (IQ):

تعد نسبة الذكاء من معايير الاختبارات شائعة الاستخدام، كما أنها أكثر المعايير غموضاً في فهمها وتفسيرها. وقد اقترح المفهوم الأصلي لنسبة الذكاء عالم النفس الألماني ويلهلم شترن Wilhelm Stern في مستهل القرن الماضي. ويرجع الفضل إلى تيرمان Terman عالم النفس الأمريكي في انتشار هذا المفهوم في الطبعة الأولى لاختبار ستانفورد بينيه للذكاء عام 1916 . ونسبة الذكاء (IQ) من هذا المنظور يُعبر عنها بالنسبة بين العمر العقلي للفرد مقسوماً على عمره الزمني، وضرب الناتج في (100) للتخلص من الكسور، أي أن:

$$\text{نسبة الذكاء (IQ)} = \frac{\text{العمر العقلي}}{\text{العمر الزمني}} \times 100 \quad \ldots\ldots\ldots\ldots\ (3-3)$$

وتعد نسبة الذكاء بمثابة مؤشر لمعدل النمو العقلي العام. غير أن العمر العقلي Mental Age لا يعطي صورة مكتملة عن النمو العقلي، حيث إن العمر العقلي (12) مثلاً يكون له تضمينات مختلفة إذا كان الطفل عمره الزمني (8) أو (14) عاماً. لذلك اقترح ضرورة قياس كل من معدل النمو العقلي، ومستوى النمو (أي العمر العقلي).

فالطفل الذي يكون ذكاؤه مناظراً لمتوسط ذكاء أطفال هذا العمر تكون نسبة ذكاؤه (IQ) تسأوي 100، وذلك لأن عمره العقلي يكون مسأوياً لعمره الزمني. أما الأطفال الذين يكون معدل نموهم العقلي أسرع من المتوسط، سوف تكون نسبة الذكاء لديهم اكبر من 100، بينما تكون أقل من 100 لمن كان معدل نموهم العقلي أبطأ من المتوسط.

والعمر العقلي الذي يُستند إليه في حساب نسبة الذكاء هو درجة خام محولة.

غير أن نسبة الذكاء يشوبها كثيراً من أوجه القصور من بينها ما يلي:

(1) الإنحرافات المعيارية لنسب الذكاء ليست ثابتة للأعمار المختلفة.

(2) القيمة القصوى للمقام في نسبة الذكاء (أي العمر الزمني) لا تزال مثار خلاف بين علماء النفس وخبراء القياس. وهذا يتعلق بتحديد العمر الذي يتوقف عنده النمو العقلي للفرد. لذلك يرى كثير من هؤلاء أن نسبة الذكاء قليلة الفائدة في قياس ذكاء الراشدين.

(3) تقلص وحدات العمر العقلي بتقدم العمر، وعلاقة ذلك بالإنحراف المعياري. فمثلاً إذا نظرنا لنسب الذكاء الآتية:

$$\left(133 = 100 \times \frac{12}{9}\right) ، \left(133 = 100 \times \frac{8}{6}\right) ، \left(133 = 100 \times \frac{4}{3}\right)$$

نلاحظ أن الطفل الذي عمره العقلي (4) يزيد عن عمره الزمني (3) بقدر عام واحد، ونسبة ذكاؤه 133، بينما الطفل الذي عمره العقلي (8) يزيد عن عمره الزمني (6) بقدر عامين، ومع هذا تظل نسبة الذكاء (133)، والطفل الذي عمره الزمني يفوق عمره العقلي بثلاثة أعوام تكون نسبة ذكاؤه 133 أيضاً. فإذا كانت نسبة الذكاء (133) تعني الشيء نفسه في جميع الأعمار،

فإن مقدار الانحراف المعياري ينبغي أن يزداد في كل عام تال، غير أن هذا ليس متحققاً في الواقع. فالنمو العقلي لا يزداد خطياً بزيادة العمر. لذلك فإن نسبة الذكاء لم تعد تُستخدم في كثير من اختبارات الذكاء الرئيسة، وإنما استعيض عنها بمعيار يعتمد على إيجاد الدرجات المعيارية المستمدة من عينة ممثلة من أفراد مجتمع لكل مستوى عمري معين. وتسمى هذه الدرجات "نسبة الذكاء الإنحرافية".

نسبة الذكاء الانحرافية Deviation IQ (DIQ):

تعد نسبة الذكاء الانحرافية نوعاً من الدرجات المعيارية المحوّلة مثل الدرجات التائية، وليست نسبة بالمفهوم الرياضي. وعادة يكون متوسطها 100 وانحرافها المعياري (15) أو (16)، على الرغم من أن بعض اختبارات الذكاء يكون الانحراف المعياري لدرجاتها (12) أو (20). ويتم حساب نسبة الذكاء الانحرافية لكل مجموعة عمرية في عينة التقنين.

فنسبة الذكاء الانحرافية (130) تضع الفرد دائماً في موقع يزيد انحرافين معياريين عن المتوسط، وهذا يناظر المئيني (98). وقد استُخدمت نسبة الذكاء (IQ) في اختبار بينيه Binet إلى أن اجريت مراجعة للاختبار عام 1960، واستُخدمت فيه عندئذ نسبة الذكاء الانحرافية (DIQ). كذلك استخدمها ويكسلر Wechsler في اختباراته المعروفة التي تقيس ذكاء الراشدين، والأطفال، وأطفال ما قبل الحضانة، والتي سوف أوضحها في الفصل السادس.

ومن الجدير بالذكر أن نسبة الذكاء، ونسبة الذكاء الانحرافية، يعدان من المعايير الارتقائية، أو النمائية Developmental Scales.

معايير الفرق الدراسية Grade Norms :

تعد معايير الفرق الدراسية أكثر المعايير استخداماً، وذلك لسهولة فهمها بواسطة مستخدمي الاختبارات. فمعيار الفرقة الدراسية Grade Norm أو معيار مكافىء الفرقة Grade Equivalent هو الدرجة المتوسطة التي يحصل عليها التلاميذ في فرقة دراسية معينة في اختبار معين. فإذا استطاع تلاميذ الفرقة الثالثة الابتدائية في الجماعة المرجعية أو عينة التقنين الإجابة الصحيحة على عدد متوسط من الأسئلة في اختبار حساب، وليكن (18)، فإن الدرجة الخام (18) تناظر الفرقة الرابعة. وتُستخدم معايير الفرق الدراسية عادة في الاختبارات التحصيلية المقننة للمرحلة الابتدائية. ودرجة الفرقة المكافئة تتحدد بكسر عشري

مثل (2.9) أو (4.4) حيث يشير الرقم الصحيح إلى الفرقة الدراسية، والجزء الكسري العشري إلى أحد شهور العام الدراسي لهذه الفرقة.

لذلك فإن الدرجة المكافئة للفرقة، وهي (2.9) تشير إلى الفرقة الثانية والشهر التاسع. وهذا يعني انه اذا كانت الدرجة (24) في اختبار تحصيلي معين تناظر الفرقة المكافئة (6.2) فإن هذا يمكن تفسيره بأن المتوسط لجميع التلاميذ الذين طبق عليهم هذا الاختبار خلال الشهر الثاني في الصف السادس كان مسأوياً (24).

ومكافىء الفرق الفعلي يمكن الحصول عليه للنقط الزمنية التي يطبق فيها الاختبار، أما مكافىء الفرق لنقط زمنية أخرى، فإنه يتم الحصول عليه بواسطة عملية استكمال Exterpolation، أو استيفاء Interpolation، وذلك بتحديد وسيط مكافئ الفرق وقت الاختبار لكل درجة من درجات الاختبار، وتوصيل هذه النقط بخط يستخدم في الاستكمال أو الاستيفاء، وبذلك يمكن تقدير القيم التي تقع بين مكافئات الفرق.

وعلى الرغم من شيوع استخدام هذا النوع من المعايير، إلا أنه يشوبه أوجه قصور متعددة من أهمها ما يلي:

(1) نمو الخصائص النفسية والتربوية لدى التلاميذ لا يكون منتظماً على المدى الكلي للفرق الدراسية. ولهذا فإن فرق شهرين من النمو في التحصيل في الفرقة الخامسة، أي من (5.3) إلى (5.5) مثلاً، لا يدل على نفس درجة التغيير في التحصيل كما في الفرقة السابعة، أي من (7.3) إلى (7.5) مثلاً. والحقيقة أن وحدة الفرقة تصبح اصغر تدريجياً كلما اتجهنا نحو الفرق الدراسية الأعلى. أي أن وحدة القياس في معايير الفرق الدراسية غير متسأوٍ كما هو الحال في معيار نسبة الذكاء التي سبق توضيحها.

(2) معايير الفرق الدراسية لا تقدم معلومات كافية فيما يتعلق بالمئيني المناظر لدرجة الفرد بالنسبة لصفه الدراسي.

(3) اختلاف تحصيل التلاميذ الذين يحصلون على مكافئات فرق متسأوية؛ فتلميذ الفرقة الرابعة الذي يكون مكافىء فرقته (7.0) لا يمتلك نفس معارف أو مهارات تلميذ الفرقة التاسعة الذي يكون مكافىء فرقته (7.5) أيضاً.

(4) معايير مكافىء الفرق الدراسية غير متقارنة عبر مجموعات فرعية من الاختبارات التي

تشتمل عليها بطارية اختبارات، وذلك للتلاميذ الذين يفوقون متوسط فرقتهم أو يقلون عنه.

الصفحات النفسية (البروفيلات) Profiles:

عندما نود عرض وتفسير درجتين أو أكثر للفرد نفسه أو لمجموعة من الأفراد في بطارية من الاختبارات، فإن هذا يتطلب استخدام ما يسمى بالصفحة النفسية أو البروفيل Profile. فالمقارنة بين هذه الدرجات لا يكون لها معنى إلا إذا أجرينا تحويلاً للدرجات الخام إلى نوع واحد من الدرجات المحوّلة استناداً إلى المجموعة المرجعية، أو عينة التقنين ذاتها. وعندما يشتمل الاختبار على عدة اختبارات فرعية مثل: اختبارات الاستعدادات متعدد العوامل، وبطاريات الاختبارات التحصيلية المقننة، وكثير من مقاييس الشخصية والميول، فإنه ينبغي تفسير درجاتها استناداً إلى معايير مستمدة من نفس العينة. وبذلك يمكن تكوين صفحات نفسية ذات معنى للأفراد المختبرين وكذلك يمكن تحديد مقدار الخطأ في القياس.

وينبغي أن تشتمل الصفحات النفسية على معلومات كافية عن الاختبار، مثل: عنوانه، وصيغته، ومستواه، واسم الفرد المختبر، وتاريخ اجراء الاختبار، والدرجات الخام التي حصل عليها الفرد في كل اختبار فرعي والتي اشتقت منها الدرجات المحولة (المعايير). وعادة تُقدم تقارير وصفية إلى جانب هذه المعايير أو الدرجات المحولة، والصفحات النفسية. وتكون هذه التقارير مطبوعة باستخدام الحاسوب كما هو الحال في اختبارات أيوا للمهارات الاساسية Iowa Tests of Basic Skills، وهي اختبارات تقيس المهارات الأساسية اللغوية والرياضية لطلاب المدارس في الولايات المتحدة الامريكية. وهذه التقارير تقدم معلومات بطريقة بسيطة يسهل فهمها بواسطة الطلاب وأولياء أمورهم.

والخلاصة أن هناك معايير متعددة ومتنوعة يُستند إليها في تفسير درجات الاختبارات والمقاييس التربوية والنفسية، ولكل منها مزاياه وعيوبه وحدود استخدامه، وكل منهما يناسب غرضاً معيناً. ومعظم مؤسسات نشر الاختبارات المقننة يستخدم جميع هذه المعايير. وبوجه عام، فإن معايير المئينيات، والتساعيات المعيارية أكثر ملاءمة للمقارنة بين درجات اختبارات فرعية Subtests. أما الدرجات المعيارية فإنها تستخدم بدرجة أكبر في المقارنة بين درجات مستويات وصيغ مختلفة لنفس الاختبار الفرعي. ويُفضل استخدام معايير الفرق الدراسية

لتقرير متوسطات مجموعات من التلاميذ إضافة إلى المعايير المئينية. وعموماً فإن تفسير الدرجات الخام للاختبارات المقننة بطرق مختلفة يثري هذا التفسير ويزيد من المعلومات المستمدة من الاختبار.

الفصل الرابع

الخصائص الأساسية للاختبارات والمقاييس

- مقدمة
- (أولاً) مفهوم الثبات وأهميته
- مصادر أخطاء القياس التي تؤثر في الثبات
- طرق تقدير ثبات درجات الاختبارات
- الخطأ المعياري للقياس
- بعض العوامل التي تؤثر في تفسير قيم معامل الثبات
- (ثانياً) مفهوم الصدق وأهميته
- أنواع صدق الاختبار
- تعقيب عام على أنواع صدق الاختبارات والمقاييس
- (ثالثاً) تحليل مفردات الاختبار وأهميته
- معامل صعوبة المفردة
- معامل تمييز المفردة
- تفسير تمييز المفردة

مقدمة:

أوضحنا في الفصل الأول أن أي أداة قياس ينبغي أن تتوافر فيها بعض الخصائص الفنية الأساسية، لعل من أهمها الثبات والصدق. وهما مفهومان مترابطان. فالاختبارات والمقاييس التربوية والنفسية ينبغي أن تُستخدم استخداماً ذكياً، حيث إن بعض الاختبارات ربما لا تكون قد أعدت إعداداً جيداً، وهنا ينبغي الحيطة أو الحذر من استخدام مثل هذه الاختبارات. وهذا الأمر لا يقتصر على المقاييس التربوية والنفسية، وإنما يمتد ليشمل أيضاً المقاييس في العلوم الفيزيائية. فما بالك بميزان أو أميتر أو ميكرومتر تعطي قراءات مختلفة إذا تكرر استخدامها في قياس الشيء نفسه في أوقات مختلفة، أليس هذا ممكنا؟ بل ويحدث كثيراً في حياتنا اليومية.

فإذا أعدت قياس وزنك مثلاً في فترات متقاربة جداً بميزان معين، وأعطاك قياساً مختلفاً، فما معنى ذلك؟

ربما تستنتج أن الميزان به خلل معين، أو أن السطح الذي يرتكز عليه الميزان ليس ثابتاً أو مستوياً، أو غير ذلك. فهذه العوامل تجعل القياس الناتج مشوباً بالخطأ. وكلما زادت هذه العوامل ازداد تأثيرها، وقلت دقة القياس، والعكس صحيح.

فالثبات اذن يعد خاصية من خصائص القياس الناتج، وليس خاصية من خصائص أداة القياس ذاتها.

فأدوات القياس ليست ثابتة أو غير ثابتة، وانما القياس الناتج عنها هو الذي يتميز بدرجة معينة من الثبات.

غير أن المشكلة تكون أكثر إلحاحاً وتعقيداً في العلوم السلوكية؛ فنتائج الاختبار ينبغي أن تكون ثابتة أو منسقة ولا تتغير تغيراً جوهرياً من تطبيق إلى آخر على الفرد نفسه ما دامت الظروف لم تتغير تغيراً ملحوظاً إذا كان لنا أن نستخدم نتائج الاختبار في اتخاذ قرارات صائبة تتعلق بالفرد أو الجماعة. فمفهوم الثبات Reliability إذن يعد من المفاهيم الاساسية في القياس النفسي والتربوي.

كذلك ينبغي أن تتميز نتائج الاختبار بالصدق Validity، أي ترتبط نتائجه بشيء خارج نطاق الاختبار ذاته. فالاختبار الذي ترتبط درجاته بالسلوك الذكي يعد مقياساً للذكاء وليس مقياساً لزمن الرجع مثلاً. والحقيقة أن مفهوم الصدق يعد من المفاهيم التي اهتم علماء القياس بدراستها دراسة مستفيضة نظراً لأهميته في اتخاذ القرارات المتعلقة بالأفراد والجماعات.

لذلك سوف أتناول في هذا الفصل كلاً من مفهومي الثبات والصدق، وأنواعهما، والطرق المختلفة لتقدير كل من هذه الأنواع والعوامل التي تؤثر فيها. وكذلك اتناول كيفية تحسين نوعية الاختبارات عن طريق تحليل مفرداتها لتحديد درجة صعوبة كل مفردة، وقدرتها على التمييز بين الأفراد في السمة المراد قياسها.

(أولاً) مفهوم الثبات وأهميته:

يشير مفهوم الثبات إلى اتساق درجات الاختبارات والمقاييس لمجموعة معينة من الأفراد، أي الاتساق عبر الزمن، أو اتساق صيغ مختلفة من نفس الاختبار، أو اتساق مفردات الاختبار ذاته، أو الاتساق عبر أفراد مختبرين ومصححين مختلفين؛ فدرجات الاختبارات ربما لا تكون متسقة لأسباب متعددة. إذ ربما يكون سلوك الأفراد غير مستقر، أو أن يكون هناك تباين في التنبؤ بهذا السلوك من وقت إلى آخر، أو اختلاف في عينة مفردات الاختبار، أو محتوى اساليب الملاحظة، أو عدم اتساق الأفراد أو إجراءات تصحيح الاختبار أو تقدير درجاته، وغير ذلك. لذلك ينبغي عند تفسير درجات الاختبار معرفة المصادر أو العوامل التي تسبب عدم اتساق هذه الدرجات. فالقياس الذي يتميز بالاتساق يعد شرطاً ضرورياً للاختبار أو المقياس الجيد (Nitko, 1983). فإذا كانت درجات الاختبار غير متسقة وتتذبذب تذبذباً كبيراً من عينة من المفردات إلى عينة اخرى، فإنه لا يكون هناك اساساً يستند اليه في اتخاذ قرار فيما يتعلق بمكانة الفرد أو تحقيقه لما يشتمل عليه النطاق السلوكي لهذه المفردات. لذلك فإن ثبات درجات الاختبارات يؤثر في نوعية القرارات المتعلقة بالأفراد أو الجماعات.

مصادر أخطاء القياس التي تؤثر في الثبات:

تتميز درجات الاختبارات بقدر مرتفع من الثبات إذا كانت تعكس الجوانب الحقيقية للسمة أو الخاصة أو القدرة المراد قياسها، ولا تعكس عوامل الخطأ والعشوائية، ويزداد ثبات

الدرجات إذا قلت هذه العوامل. وتوجد مصادر متعددة تسهم في أخطاء القياس التي تؤدي إلى انخفاض ثبات درجات الاختبارات والمقاييس التربوية والنفسية، بعضها يتعلق بأداة القياس ذاتها، وبعضها يتعلق بإجراءات وظروف تطبيقها، وبعض آخر يتعلق بالأفراد المختبرين.

أخطاء تتعلق بأداة القياس:

سبق أن اوضحنا في الفصل الأول أن هناك خصائص عامة ينبغي توافرها في أدوات القياس لكي تكون درجاتها متسقة بدرجة كبيرة. فعدم ملاءمة محتوى الاختبار للأفراد المختبرين، ومستوى صعوبة مفرداته، أو غموضها وعدم دقتها، وكذلك غموض تعليمات الإجابة عليها، وطول الاختبار، والزمن اللازم للإجابة عليه، واشتماله على عينة غير ممثلة لمكونات السمة المراد قياسها، وغير ذلك تعد من مصادر الخطأ المتعلقة بأداة القياس، وتؤدي إلى عدم اتساق الدرجات، كما تؤثر أيضاً في صدق الاختبار، حيث إن المعلومات والقرارات التي تستند إليها تكون غير موثوق بها، أي غير صادقة.

أخطاء تتعلق بإجراء الاختبار، وظروف تطبيقه:

يتأثر ثبات درجات الاختبارات ليس فقط بتصميم وبناء أداة القياس التي تستخدم في جمع المعلومات، وإنما ايضاً بكيفية استخدام الأداة. فتطبيق الاختبارات يتطلب أن تكون تعليمات الإجابة واضحة وكافية ومناسبة للمختبرين. كما أن ظروف تطبيق الاختبار ينبغي أن تكون جيدة ومقننة، فالاختبار الذي يطبق على مجموعة كبيرة من الأفراد في غرفة ضيقة وغير جيدة التهوية، وقليلة الاضاءة، وتحيط بها الضوضاء تؤثر تأثيراً كبيراً في درجات الأفراد، مما يؤدي إلى انخفاض ثبات هذه الدرجات وعدم صدق المعلومات المستمدة منها. وكذلك تعد عملية وإجراءات تصحيح الاختبار أو تقدير درجاته، وتدوين نتائجه من مصادر الأخطاء الشائعة. فمن السهل الوقوع في الأخطاء الحسابية عند جمع الدرجات، أو إغفال تصحيح بعض الأسئلة، أو عدم الدقة في تصحيح الأسئلة التي تتطلب استجابات مفتوحة، وغير ذلك.

أخطاء تتعلق بالأفراد المختبرين:

توجد أخطاء متعددة تتعلق بالأفراد المختبرين، وتسهم في انخفاض ثبات درجات الاختبارات. ومن أمثلة ذلك:

انخفاض الدافعية لدى المختبرين وقت تطبيق الاختبار لأسباب ربما ترجع إلى عدم معرفتهم بالغرض من الاختبار أو اتجاههم السلبي نحوه، أو قلة انتباههم، أو شعورهم بالإجهاد أو التعب أو الملل، أو انخفاض قدرتهم القرائية، أو حالتهم الصحية والنفسية، وايضاً مستواهم الاجتماعي الاقتصادي، أو تفاوت تدريبهم على نوع معين من الاختبارات. كما أن عدم استقرار بعض السمات الإنسانية وتذبذبها عبر الزمن يجعل من الضروري الحصول على معلومات تتعلق بهذه السمات في الوقت الذي سوف تستخدم فيه هذه المعلومات في اتخاذ قرارات بشأن الأفراد المختبرين.

لذلك ينبغي معرفة مصادر الخطأ الممكنة التي يمكن أن تؤثر في استجابات الأفراد للاختبارات والمقاييس التربوية والنفسية، وتؤدي إلى عدم اتساق الدرجات، مما يسهم في تحسين طرق القياس، واستخدام نتائج القياس استخداماً مستبصراً، ومعرفة حدود الدقة في تفسير هذه النتائج.

طرق تقدير ثبات درجات الاختبارات:

لكي تكون هناك ثقة في استخدام نتائج الاختبارات والافادة منها ينبغي ضبط مصادر الأخطاء المتعددة التي أوضحناها. فتقنين الاختبار، أي ضبط جميع العوامل المغترية بقدر الامكان يؤدي إلى زيادة ثبات الدرجات المستمدة منه.

والدرجات التي يحصل عليها الفرد في الاختبار تسمى الدرجات الملاحظة Obtained Scores، وسواء كانت هذه الدرجات تقديرات أو درجات خام، أو درجات محولة، فإنه يشوبها جميعاً قدراً من الخطأ. فإذا استطعنا تقدير هذا الخطأ في الدرجات الملاحظة فإنه يكون لدينا "درجة الخطأ Error Score"، ويمكن الحصول على "الدرجة الحقيقية True Score" للفرد بطرح درجة الخطأ من درجته الملاحظة.

أي أن الدرجة الحقيقية = الدرجة الملاحظة - درجة الخطأ.

وهذا يعني أن الدرجة الملاحظة تتكون من الدرجة الحقيقية، ودرجة الخطأ، ويطلق على درجة الخطأ "خطأ القياس Error of Measurement"، ودرجة الخطأ يمكن أن تكون موجبة أو سالبة أو صفر. فإذا كانت درجة الخطأ موجبة تزيد الدرجة الملاحظة عن الدرجة الحقيقية، والعكس إذا كانت درجة الخطأ سالبة.

ومعامل الثبات Reliability Coefficient يزودنا بتقدير كمي للعلاقة بين الفروق الفردية في الدرجات الحقيقية، والفروق الفردية في الدرجات الملاحظة. والدرجات الملاحظة الخالية من الأخطاء تنتج عن اختبار تتسم درجاته بالثبات التام. وهذا يجعل الفروق الفردية في الدرجات الملاحظة مساوياً للفروق الفردية في الدرجات الحقيقية، وتكون قيمة معامل الثبات في هذه الحالة مساوياً الواحد الصحيح. أما الاختبار الذي تتسم درجاته بعدم الثبات على الإطلاق فسوف تكون الدرجات الملاحظة المستمدة منه ليست إلا اخطاء قياس، وعندئذ تكون قيمة معامل الثبات صفراً.

لذلك فإن قيمة معامل الثبات تتراوح بين (صفر ، 1)، وهو بذلك يعكس درجة اتساق درجات الاختبار، ويمكن تفسيره بأنه نسبة من الفروق الملاحظة بين الأفراد التي تعزى إلى الفروق الحقيقية بينهم.

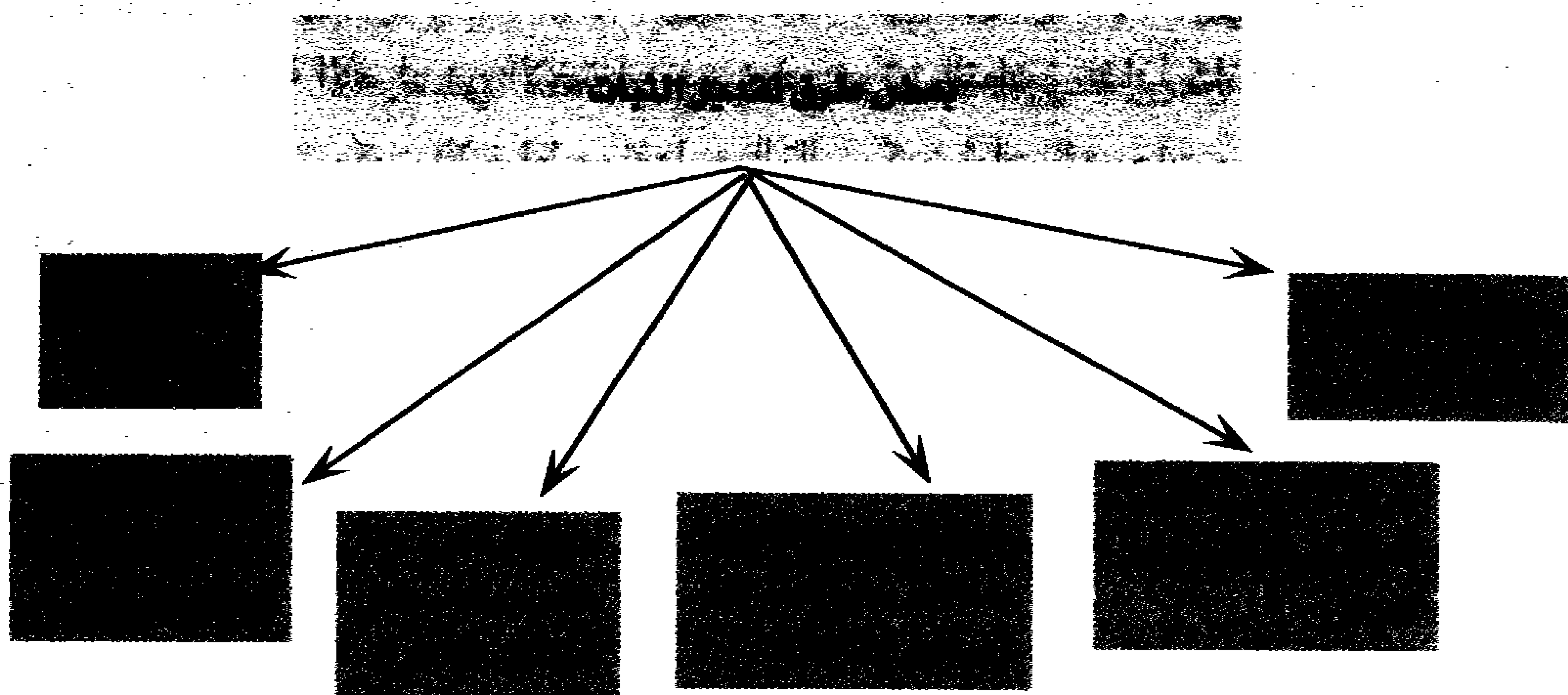
ويمكن تكميم الفروق في الدرجات الحقيقية بين الأفراد باستخدام تباين هذه الدرجات، وكذلك الفروق في الدرجات الملاحظة.

لذلك يمكن تعريف معامل الثبات بأنه النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة.

$$\text{أي أن معامل الثبات} = \frac{\text{تباين الدرجات الحقيقية}}{\text{تباين الدرجات الملاحظة}}$$

وهذا يعني أن معامل الثبات هو مدى ما تعكسه الفروق في الدرجات الملاحظة على الفروق في الدرجات الحقيقية، أو مدى إعزاء تباين الدرجات الملاحظة إلى تباين الدرجات الحقيقية. ونظراً لأن تباين الدرجات الحقيقية يكون غير معلوم، فإننا نود تقدير الدرجة الحقيقية للفرد باستخدام درجته الملاحظة، وهذا يتطلب تقدير درجة الخطأ أو تباين الخطأ، وهو ما يهدف إليه مفهوم الثبات.

وتختلف طرق تقدير ثبات درجات الاختبارات باختلاف مصادر الخطأ الذي نود التحكم فيها بحيث لا ينعكس أثرها في الدرجة الملاحظة. ويوضح الشكل التخطيطي التالي (4 - 1) أهم هذه الطرق شائعة الاستخدام في الاختبارات المقننة المنشورة:



• شكل (4 - 1) يوضح بعض طرق تقدير الثبات

وفيما يلي توضيحاً لكل من هذه الطرق والاستخدام المناسب لها:

(1) طريقة إعادة الاختبار (معامل الاستقرار) Test - retest (Stability) :

للتحقق من استقرار درجات الاختبار، يطبق الاختبار على عينة ممثلة من الأفراد، ثم يعاد تطبيقه بعد مدة زمنية معينة. وتتوقف هذه المدة الزمنية الفاصلة على طبيعة السمة التي يقيسها الاختبار. فإذا كان يقيس سمة سريعة التذبذب والتغير، مثل: ذكاء الأطفال الصغار، أو السمات المزاجية، فإن هذه الطريقة لا تكون مناسبة، ويحسن استخدام طريقة أخرى من الطرق التي سنذكرها. وإذا كانت المدة الفاصلة قصيرة، فإن الذاكرة قصيرة الأمد سوف تؤثر في نتائج التطبيق في المرة الثانية بحيث تجعلها أكثر اتساقاً مع نتائج التطبيق الأول. وإذا كانت هذه المدة طويلة، فإن النضج، والعوامل البيئية، سوف تؤثر أيضاً في نتائج التطبيق الثاني. وعندئذ يصبح معامل الثبات أو ما يسمى في هذه الحالة "معامل الاستقرار Stability Coefficient" محل تساؤل. لذلك يجب العناية باختيار المدة الزمنية الفاصلة اعتماداً على طبيعة السمة المراد قياسها، وخصائص المجتمع المستهدف.

ويمكن توضيح هذه الطريقة بالشكل (4 - 2) التالي:



شكل (4 - 2) يوضح طريقة اعادة الاختبار

ويمكن عندئذ إيجاد معامل ارتباط بيرسون الذي سبق توضيحه في الفصل الثالث بين درجات مجموعة الأفراد في الاختبار نفسه في مرتي التطبيق. فإذا كانت استجابة الأفراد متسقة في المرتين، يكون الارتباط مرتفعاً، وبالتالي يكون ثبات الدرجات مرتفعاً. ويحسن ذكر المدة الزمنية الفاصلة إلى جانب قيمة معامل الثبات التي نحصل عليها.

ومن الجدير بالذكر أن التغيير الذي يحدث في الدرجات في المرتين يُنظر اليه على انه نتيجة أخطاء القياس، بافتراض أن السمة ليست سريعة التغير، أي مستقرة نسبياً.

(2) طريقة تطبيق صيغتين متكافئتين للاختبار (معامل التكافؤ Equivalence Coefficient):

يصعب أحياناً تطبيق الاختبار على نفس المجموعة من الأفراد مرتين، أو أن السمة المراد قياسها تتميز بعدم الاستقرار، فعندئذ يمكن استخدام صيغتين متكافئتين للاختبار نفسه، وتطبيقهما على مجموعة الأفراد نفسها الواحدة تلو الأخرى، كما يوضحه الشكل (4 - 3) التالي:



شكل (4 - 3): يوضح طريقة تطبيق صيغتين متكافئتين للاختبار

وتستند هذه الطريقة على فرضية أن كلاً من صيغتي الاختبار تكونان في الحقيقة متكافئتين من حيث المحتوى، ونوع المفردات، وعددها، وصعوبتها ، والمتوسط الانحراف المعياري لدرجات كل منهما.

ويمكن إيجاد معامل الارتباط بين درجات كل من الصيغتين للحصول على معامل الثبات الذي يسمى في هذه الحالة "معامل التكافؤ Equivalence Coefficient".

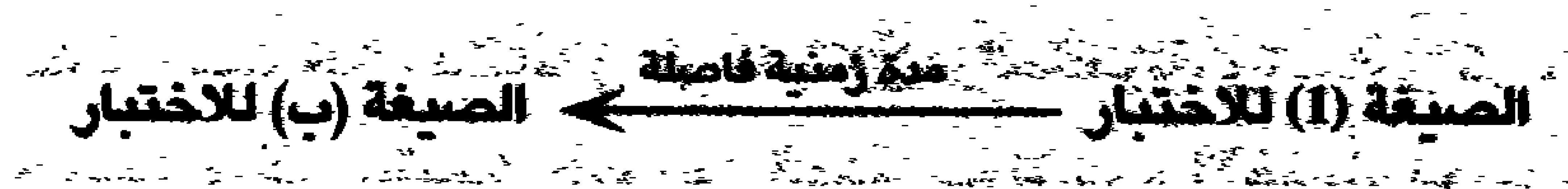
وترجع أخطاء القياس في هذه الحالة إلى اختلاف مفردات أو اسئلة كل من الصيغتين؛ فكلما زاد الاختلاف بينهما انخفض الثبات. أي أن أخطاء القياس في هذه الحالة ترجع إلى اختلاف عينات الاسئلة في كل من الصيغتين، وليس إلى التغيرات التي تحدث للأفراد كما هو الحال في معامل الاستقرار.

وكثير من الاختبارات المقننة يكون لها صيغتين متكافئتين على الأقل للاستخدام في حالة

تغيب بعض الأفراد عن الاختبار، أو لأي أسباب أخرى تتطلب عدم تذكر الإجابات في الصيغة الأولى، أو غير ذلك. وتدوّن تقديرات معامل التكافؤ في ادلة هذه الاختبارات.

(3) طريقة تطبيق صيغتين متكافئتين بفاصل زمني: (Stability & Equivalence)

أحياناً نهتم بالتنبؤ بعيد المدى، والاستدلال على اداء الفرد في نطاق من المعارف أو المهارات، فعندئذ نحتاج إلى معامل يشمل كل من معاملي الاستقرار والتكافؤ، فإذا كان لدينا صيغتين للاختبار، فإننا نطبق الصيغة الأولى، وبعد انقضاء مدة زمنية معينة نطبق الصيغة الثانية كما هو موضح بالشكل (4 - 4) التالي:



شكل (4 - 4): يوضح طريقة تطبيق صيغتين متكافئتين بفاصل زمني

ويمكن ايجاد معامل الارتباط بين مجموعتي الدرجات في كل من صيغتي الاختبار.

ويلاحظ أن جميع مصادر الخطأ التي تؤثر في الطريقتين الأولى والثانية والتي تؤدي إلى عدم اتساق درجات الاختبار تعد مصادر أخطاء في هذه الطريقة. لذلك يتوقع انخفاض قيمة معامل الثبات في هذه الحالة عنها في الطريقتين السابقتين. ويسمى المعامل الناتج "معامل الاستقرار والتكافؤ Stability & Equivalence Coefficient".

غير أن هذا المعامل يكون ذا أهمية في قياس التحسن في الأداء، أو عندما نود التعميم على المواقف أو الظروف، وعينات مفردات الاختبار. ونعني بذلك مدى دقة الاختبار في قياس سمة معينة بغض النظر عن صيغة الاختبار المستخدمة، أو الظروف التي طبق فيها الاختبار.

ويحدث ذلك في كثير من الأحيان، فإذا أردنا تقدير قدرة فرد معين على التدريس مثلاً، فإننا ربما نود التركيز على اتساق التقديرات التي يتم الحصول عليها في ظروف ومواقف مختلفة بواسطة أكثر من شخص قائم بالتقدير. لذلك ينبغي عند استخدام هذا المعامل في تقدير ثبات درجات الاختبارات تحديد الظروف والمواقف بدقة لكي يستند إليها تفسير هذه الدرجات.

(4) طريقة التجزئة النصفية (معامل الاتساق الداخلي Internal Consistency Coefficient):

الطرق الثلاث السابقة لتقدير الثبات تناسب بدرجة كبيرة الاختبارات المقننة – Standardized. غير انه في كثير من المواقف يصعب تقدير معامل الاستقرار، أو أي من معاملي التكافؤ السابقة. فالمعلم في الصف المدرسي ربما لا يود تطبيق صيغتين للاختبار إذا توافرتا، أو يعيد تطبيق الاختبار نفسه في مرتين متتاليتين. لذلك توجد عدة طرق لتقدير الثبات تعتمد على تطبيق الاختبار مرة واحدة. وتعد طريقة التجزئة النصفية Split- Half Method أكثر هذه الطرق استخداماً. وتتطلب هذه الطريقة تقسيم الاختبار إلى نصفين متكافئين إحصائياً. وهذان النصفان يكونان مستقلين ومتكافئين في المحتوى، والصعوبة، ويتساوى كل من المتوسط، والانحراف المعياري لدرجاتهما. وليس من الضروري تقسيم الاختبار بحسب الأسئلة أو المفردات الفردية والزوجية إلا إذا حقق هذا التقسيم الشروط المذكورة.

وبعد تطبيق الاختبار كوحدة وتصحيح نتائجه، ترصد درجات أسئلة أو مفردات النصف الأول على حدة، ودرجات أو أسئلة النصف الثاني على حدة، ثم نوجد معامل الارتباط بين درجات كل من النصفين.

ونظراً لأن معامل الارتباط الذي نحصل عليه يكون مقياساً لثبات درجات نصف الاختبار، فإنه يجب استخدام القيمة الناتجة في تقدير ثبات درجات الاختبار ككل، حيث إن قيمة معامل ثبات درجات نصف الاختبار تكون أقل من ثبات درجات الاختبار التي كان من الممكن الحصول عليها من الاختبار بطوله الأصلي.

وقد اقترح سبيرمان وبراون Spearman & Brown صيغة رياضية يمكن استخدامها في تقدير معامل ثبات درجات الاختبار ككل، وذلك من قيم معامل الارتباط بين درجات كل من نصفي الاختبار. وهذه الصيغة كالتالي:

$$\text{معامل ثبات درجات الاختبار} = \frac{2 \times \text{الارتباط بين نصفي الاختبار}}{1 + \text{الارتباط بين نصفي الاختبار}}$$

$$ر_{11} = \frac{2\,ر_{21}}{1 + ر_{21}} \qquad \text{.......... (1 - 4)}$$

حيث ترمز ($ر_{11}$) إلى معامل ثبات درجات الاختبار الذي يتم تقديره.

، ($ر_{21}$) إلى معامل الثبات الناتج عن التجزئة النصفية.

فإذا كان معامل الارتباط بين نصفي الاختبار ($ر_{21}$ = 0.75)،

$$\text{فإن تقدير معامل ثبات الاختبار ككل} = \frac{0.75 \times 2}{1 + 0.75} = \frac{1.50}{1.75} = 0.86 \text{ تقريباً}$$

ويلاحظ أن قيمة ثبات درجات الاختبار ككل قد زادت من 0.75 إلى 0.86، وهذا يدل على أن ثبات درجات الاختبار يزداد بازدياد طول الاختبار بشرط أن تقيس أسئلته أو مفرداته التي أدت إلى زيادة الطول السمة نفسها التي تقيسها بقية الأسئلة أو المفردات.

ويمكن تعميم الصيغة (4 - 1) لتقدير ثبات درجات الاختبار إذا زاد أو نقص طوله (ن) من المرات لتصبح كالتالي:

$$ر_{ن ن} = \frac{ن\ ر_{ك ك}}{1 + (ن - 1)\ ر_{ك ك}} \quad \text{.......... (4 - 2)}$$

حيث ترمز ($ر_{ن ن}$) إلى القيمة التقديرية لمعامل الثبات إذا زاد طوله (ن) من المرات.

(ن) تساوي عدد مفردات الاختبار بعد الإطالة مقسوماً على عدد مفرداته الأصلية.

، ($ر_{ك ك}$) إلى ثبات درجات الاختبار الأصلي قبل الإطالة.

ويلاحظ اننا إذا وضعنا ن = 2 في الصيغة (4 - 2) فإننا نحصل على الصيغة (4 - 1) السابقة التي تعد حالة خاصة منها. ويتطلب استخدام الصيغة العامة (4 - 2) تحقق الشروط التي سبق ذكرها.

فإذا كان معامل ثبات درجات الاختبار الأصلي 0.75، وزاد طوله ثلاث مرات، فإن القيمة التقديرية لثبات الدرجات بعد إطالة الاختبار تكون كالتالي:

$$ر = \frac{0.75 \times 3}{1 + 0.75\ (1 - 3)} = \frac{2.25}{2.50} = 0.90\ .$$

وينبغي التحقق من تساوي تباين درجات كل من نصفي الاختبار. اما إذا اختلف التباين، فإنه يمكن استخدام الصيغة التي اقترحها جاتمان (Guttman, 1945) وهي كالتالي:

$$ر_{ن ن} = 2 \left[1 - \frac{ع_{أ}^{2} + ع_{ب}^{2}}{ع_{ن}^{2}} \right] \quad (3 - 4) ....$$

حيث ترمز ($ع_{أ}$) ، ($ع_{ب}$) إلى الانحراف المعياري لدرجات كل من نصفي الاختبار.

، ($ع_{ن}$) إلى الانحراف المعياري لدرجات الاختبار ككل.

وتطبيق هذه الصيغة لا يتطلب إيجاد الارتباط بين درجات نصفي الاختبار، وإنما يكتفي بإيجاد الانحرافات المعيارية للدرجات.

ويلاحظ أن أي من هذه الصيغ يتطلب تطبيق الاختبار مرة واحدة، أي تطبيق كل من نصفي الاختبار في الوقت نفسه. ومع هذا يعاب على طريقة التجزئة النصفية أن المصدر الرئيس للخطأ يتعلق بالمفردات ذاتها، وهذا يعني أن عدم الثبات ينتج عن الفروق في عينة المفردات التي يشتمل عليها كل من نصفي الاختبار. لذلك فإن هذه الطريقة تؤدي إلى قيم مرتفعة نسبياً لمعامل الثبات. وينبغي عدم استخدام هذه الطريقة في تقدير ثبات درجات الاختبارات التي تعتمد على السرعة، وذلك لأنها تؤدي إلى قيم زائفة لمعامل الثبات، وإنما يقتصر استخدامها على الاختبارات التي تعتمد على القوة.

(5) طريقة كيودر - ريتشاردسون (معامل التجانس) Homogeneity Coefficient:

تعتمد هذه الطريقة أيضاً على تطبيق الاختبار في مرة واحدة، ويمكن باستخدامها التحقق من تجانس جميع مفردات الاختبار Test Homogeneity، وهذا التجانس يتأثر بكل من معاينة المحتوى Content Sampling كما في طريقة التجزئة النصفية، وعدم اتساق النطاق السلوكي الذي تستمد منه عينة مفردات الاختبار. فكلما زاد هذا الاتساق دل ذلك على تجانس مفردات الاختبار. فالاختبار الذي يشتمل على مفردات تقيس عملية الجمع البسيط فقط، يكون اكثر اتساقاً أو تجانساً من الاختبار الذي يقيس عمليتي الجمع والطرح. وتتغلب هذه الطريقة على مشكلة تجزئة الاختبار إلى نصفين متكافئين، حيث إنها تعتمد على إيجاد قيمة تقديرية لمتوسط قيم معاملات الارتباط بين كل من نصفي الاختبار إذا تم تجزئته بجميع الطرق الممكنة دون أن نلجأ إلى هذه التجزئة فعلياً.

وقد اقترح كل من كيودر وريتشاردسون Kuder & Richardson صيغتين رياضيتين

أطلقا عليهما الصيغة (20) والصيغة (21)، ويرمز لهما بالرمزين (K - R 20) , (k - R 21). غير انني سأقتصر على ذكر الصيغة (20) لأنها أكثر شيوعاً واستخداماً في الاختبارات والمقاييس، وأكثر دقة من الصيغة (21) :

$$\text{معامل التجانس} = \frac{ن}{ن - 1} \times \left(\frac{ع^2 - \text{مجـ س ص}}{ع^2} \right) \quad \ldots\ldots\ (4 - 4)$$

$$= \frac{ن}{ن - 1} \times \left(1- \frac{\text{مجـ س ص}}{ع^2} \right) \quad \ldots\ldots\ (5 - 4)$$

حيث ترمز (ن) إلى العدد الكلي لمفردات الاختبار.

، (س) إلى نسبة عدد الأفراد الذين أجابوا على أي مفردة في الاختبار إجابة صحيحة.

، (ص) إلى نسبة عدد الأفراد الذين أجابوا على أي مفردة اجابة خطأ، حيث (ص = 1− س).

، (مجـ س ص) إلى مجموع تباين درجات مفردات الاختبار.

، (ع 2) إلى تباين الدرجات الكلية في الاختبار (مربع الانحراف المعياري).

ويتطلب تطبيق هذه الصيغة إيجاد قيمة الانحراف المعياري لدرجات الاختبار كما أوضحنا في الفصل الثالث، ونسبة عدد الإجابات الصحيحة والإجابات الخطأ لكل مفردة من مفردات الاختبار، وإيجاد حاصل ضرب النسبتين لكل مفردة وجمع الناتج.

فإذا كانت $ع^2 = 2.5$ ، مجـ س ص = 1.09 ، ن = 5 ،

$$\text{فإن معامل التجانس} = \frac{5}{1 - 5} \left(\frac{1.09 - 2.5}{2.5} \right) = \frac{5}{4} \times \frac{1.41}{2.5} = 0.71 .$$

وتفترض هذه الطريقة أن الاختبار من نوع القوة كما في طريقة التجزئة النصفية، لذلك فهي لا تناسب الاختبارات التي تعتمد على السرعة، كما تفترض أن مفردات الاختبار تقيس سمة واحدة أو بُعد واحد، وأن المفردات ثنائية الدرجة كما في مفردات الاختيار من متعدد، أو الصواب أو الخطأ. وتعد طريقة التجزئة النصفية حالة خاصة من هذه الطريقة.



معامل (α) لكرونباك (α) Coefficient:

نظراً لأن معامل التجانس السابق يقتصر استخدامه على المفردات ثنائية الدرجة. فقد اقترح كرونباك (Cronbach, 1951) طريقة تناسب اختبارات القوة التي تشتمل على مفردات متعددة الدرجات ، مثل: أسئلة المقال، أو فقرات الاستبانات التي تتطلب إجابتها الاختيار من بين بدائل متعددة، وتكون درجاتها متدرجة من 1 إلى 5 مثلاً.

والصيغة الرياضية لمعامل (α)، ويقرأ (معامل ألفا) كالتالي:

$$\text{معامل } \alpha = \frac{\text{ن}}{\text{ن}-1}\left[1- \frac{\text{مجـ ع}^2_{\text{هـ}}}{\text{ع}^2_{\text{ك}}}\right] \quad \text{........ (4 - 6)}$$

حيث ترمز (ن) إلى العدد الكلى لمفردات الاختبار.

، (مجـ $\text{ع}^2_{\text{هـ}}$) إلى مجموع تباين درجات كل مفردة (هـ).

، ($\text{ع}^2_{\text{ك}}$) إلى تباين الدرجات الكلية في الاختبار.

والصيغة (4- 6) تماثل الصيغة (4 - 5) السابقة فيما عدا الرمز (مجـ $\text{ع}^2_{\text{هـ}}$) الذي يمثل تباين درجات جميع الأفراد في إحدى المفردات التي يقوم بتقدير درجاتها محكم واحد. ومن ثم نجمع تباين درجات جميع المفردات.

فإذت كانت ن = 4 مفردات.

، تباين درجات المفردات الأربع ($\text{ع}^2_{\text{هـ}}$) = 1.81 + 1.00 + 2.00 + 1.58 = 6.39 .

، وتباين الدرجات الكلية في الاختبار يتطلب جمع درجات كل فرد في المفردات الأربع، وحساب الانحراف المعيارى لهذه الدرجات الكلية كما سبق أن أوضحنا في الفصل الثالث، وتربيع القيمة الناتجة للحصول على التباين وهو يساوي 16.89.

ولإيجاد قيمة معامل (α) نستخدم الصيغة (4 - 6) كالتالي:

$$\text{معامل } (\alpha) = \frac{4}{4-1}\left(1- \frac{6.39}{16.89}\right) = \frac{4}{3}(1 - 0.38)$$

$$= (1.33)(0.62) = 0.83$$

ومن الجدير بالذكر أن صيغتي كيوردر - ريتشاردسون، وكرونباك تستندان إلى أن اتساق استجابات الأفراد عبر مفردات الاختبار يمكن الاعتماد عليه في تقدير معامل الثبات للدرجات الكلية في الاختبار. فإذا كانت مفردات الاختبار متجانسة، أى تقيس سمة أو بُعد واحد، فإن كلاً من الصيغتين تؤدي إلى نتيجة قريبة من النتيجة التي نحصل عليها من تطبيق طريقة التجزئة النصفية. أما إذا كانت المفردات غير متجانسة فإن قيمة معامل الثبات الناتجة عن تطبيق كل من الصيغتين سوف تكون أقل من القيمة الناتجة عن التجزئة. والحقيقة أن قيم كل من معامل التجانس (الصيغة 20)، أو معامل (α) تعد بمثابة الحد الأدنى Lower Bound للقيم التقديرية لمعامل الثبات.

الخطأ المعياري للقياس :

نظراً لأن القياس النفسي والتربوى يتأثر بمصادر أخطاء متعددة كما أوضحنا، فإنه لا يوجد اختبار يقيس سمة معينة لدى الفرد قياساً متسقاً اتساقاً تاماً. لذلك ينبغي تقدير قيمة هذه الأخطاء، وأخذها بعين الاعتبار عند تفسير درجات الاختبار. فإذا استطعنا تقدير الدرجات الحقيقية للأفراد في الاختبار، فإن الانحراف المعياري لتوزيع هذه الدرجات يُطلق عليه «الخطأ المعياري للقياس Standard Error of Measurement (SEM) وهو مؤشر لمدى اختلاف الدرجات الملاحظة للأفراد في الاختبار عن درجاتهم الحقيقية. ويفيد هذا المؤشر في تقدير قيمة الخطأ المعياري للقياس لكل فرد في الاختبار دون الحاجة إلى تكرار تطبيق الاختبار عليه عدة مرات لتعرف مدى اتساق درجاته.

والخطأ المعياري للقياس هو الانحراف المعياري للخطأ، ويزودنا بقيمة مطلقة وليست نسبية لمدى انطباق الدرجات الملاحظة والدرجات الحقيقية. فمعاملات الثبات تعتمد على تباين درجات مجموعة الأفراد، بينما الخطأ المعياري للقياس لا يتأثر بهذا التباين. وتقدير الخطأ المعياري للقياس لا يتطلب معرفة الدرجات الحقيقية.

وقد سبق أن أوضحنا في مستهل هذا الفصل أن معامل الثبات ($ر_{11}$) هو النسبة بين تباين الدرجات الحقيقية ($ع^2_ح$) ، وتباين الدرجات الملاحظة ($ع^2_م$).

أى أن:

$$ر_{11} = 1 - \frac{ع^2_ح}{ع^2_م} \quad \text{............ (7 - 4)}$$

ولكن تباين الدرجات الحقيقية = تباين الدرجات الملاحظة – تباين درجات الخطأ.

أى أن : $ع_ح^2 = ع_م^2 - ع_خ^2$ .... (4 - 8)

وبالتعويض من (4 - 8) في (4 - 7) نجد أن :

$$ر_{11} = \frac{ع_م^2 - ع_خ^2}{ع_م^2} = \frac{ع_م^2}{ع_م^2} - \frac{ع_خ^2}{ع_م^2} = 1 - \frac{ع_خ^2}{ع_م^2}$$

أى أن : $ر_{11} = 1 - \frac{ع_خ^2}{ع_م^2}$

$$1 - ر_{11} = \frac{ع_خ^2}{ع_م^2}$$

$$ع_خ^2 = ع_م^2 \left(1 - ر_{11}\right)$$

$$ع_خ = ع_م \sqrt{1 - ر_{11}}$$ .... (4 - 9)

وهذه هي الصيغة المستخدمة في إيجاد القيمة التقديرية للخطأ المعيارى للقياس، أو الانحراف المعيارى للدرجة الملاحظة للفرد في الاختبار (بالتطبيق المتكرر للاختبار نفسه) حول درجة الحقيقية في الاختبار.

فإذا افترضنا أن أخطاء القياس للفرد في التطبيقات المتكررة للاختبار تعد أخطاءً عشوائية، فإن متوسط الدرجات الملاحظة في هذه المرات يكون تقديراً لدرجته الحقيقية، وذلك لأن متوسط درجات الأخطاد العشوائية يساوي صفراً. لذلك فإن التوزيع الافتراضي للدرجات الملاحظة يكون اعتدالياً متوسطه هو الدرجة الحقيقية للفرد في الاختبار.

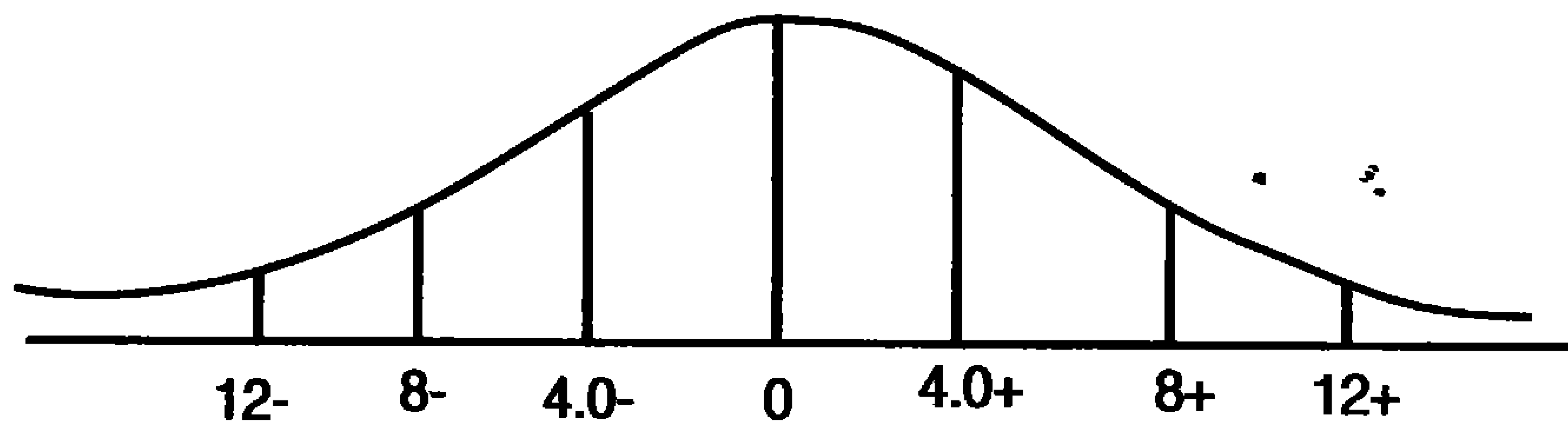
ويمكن استخدام خصائص المنحنى الاعتدالي المعياري التي سبق توضيحها في الفصل الثالث في تفسير القيمة التقديرية للخطأ المعياري للقياس التي تنتج عن تطبيق الصيغة (4- 9).

ولتوضيح ذلك نفترض أن الانحراف المعياري للدرجات الملاحظة = 10 ، ومعامل ثبات درجات الاختبار = 0.84 ، فإن الخطأ المعياري للقياس :

$$ع_{خ} = ع_{م}\sqrt{1 - ر_{11}}$$

$$= 10\sqrt{1 - 0.84} = 10\sqrt{0.16} = 10\,(0.4) = 4.0.$$

ولتفسير هذه القيمة التقديرية لأخطاء القياس نرسم المنحنى الاعتدالي، ونحدد عليها هذه القيمة باعتبارها تمثل الانحراف المعياري للأخطاء كالتالي:



شكل (4-5) يوضح التوزيع الاعتدالي لدرجات الخطأ التي
متوسطها صفر، وانحرافها المعياري = 4.0

ويتضح من شكل (4-5) أننا نتوقع أن تنحصر درجة الفرد الحقيقية بين (الدرجة الملاحظة + الخطأ المعياري للقياس)، أي بين (الدرجة الملاحظة + 4)، 68% من المرات .

فإذا كانت درجته الملاحظة 20 ، فإن درجته الحقيقيـــة يحتمل أن تنحصر بين (20– 4، 20 + 4)، أي بين (16 ، 24) عدد من المرات نسبتها 68% تقريباً. وبذلك يمكن معرفة حدود الخطأ في الدرجة التي يحصل عليها الفرد من الاختبار، وتقرير ذلك عند تفسير هذه الدرجة.

بعض العوامل التي توثر في تفسير قيم معامل الثبات:

عند تفسير القيم التقديرية لمعامل الثبات التى نحصل عليها باستخدام أى من الطرق السابقة، وبخاصة عند مقارنة معاملات ثبات درجات اختبارات مختلفة، ينبغي مراعاة بعض العوامل، نذكر منها ما يلي:

(1) الدرجات المستمدة من اختبار عدد مفرداته أو أسئلته كبيراً نسبياً تكون أكثر ثباتاً مما لو كان عدد المفردات قليلاً. ولعل هذا يتضح إذا رجعنا إلى الصيغة العامة لسبيرمان وبراون (4 - 2) المذكورة سابقاً.

(2) تتذبذب القيم العددية لمعامل الثبات من عينة إلى اخرى من الأفراد.

(3) كلما كان مدى الدرجات ضيقاً قلت قيمة معامل الثبات، وهذا يعني انه كلما كانت قدرة الأفراد المختبرين متقاربة، فان قيمة معامل الثبات تميل الي الانخفاض، وذلك لأنها تعتمد على معامل الارتباط الذي تتأثر قيمته بتشتت الدرجات كما ذكرنا في الفصل الثاني.

(4) كلما زادت المدة الزمنية الفاصلة بين مرتي تطبيق الاختبار، انخفضت قيمة كل من معامل الاستقرار، ومعامل التكافؤ.

(5) الموضوعية في تصحيح الاختبار أو تقدير درجاته تزيد من ثبات درجات الاختبار. أما الذاتية التى تعنى عدم اتساق التصحيح، فإنها تسهم في الأخطاء العشوائية التى تؤدى إلى انخفاض قيمة معامل الثبات.

(6) زيادة صعوبة الاختبار، وصعوبة مفرداته يجعل الأفراد يلجأون إلى التخمين العشوائى في إجابتهم، مما يزيد من الأخطاء العشوائية، ويقلل من ثبات درجات الاختبار.

مفهوم الصدق وأهميته:

تؤدي معظم الاختبارات التربوية والنفسية إلى نتائج كمية يُستند اليها في وصف سمات معينة لدى الأفراد. غير أننا لا نهتم كثيراً في معظم الأحيان بالدرجات ذاتها، وإنما ما يمكن أن تمثله أو تفسره هذه الدرجات. وبالطبع نود أن يكون هذا التفسير دقيقاً وموثوقاً به ثقة

كاملة، وهو ما لا يتحقق في الواقع الفعلي؛ فمثلاً إذا طبقنا اختباراً على أحد الأفراد، وأعدنا تطبيقه عليه بعد مدة زمنية معينة، فإننا لا نتوقع أن يحصل هذا الفرد على الدرجة نفسها في المرتين .

وكذلك إذا أردنا التنبؤ بأداء طلب الثانوية العامة في الكلية التى سوف يلتحق بها، فإنه يكون لدينا نتائج اختبار استعداد دراسى معلوم علاقته بالتقديرات في العام الدراسي الأول بالكلية، واستناداً إلى درجة الطالب في الاختبار نتنبأ بأنه سيحصل على تقدير معين، غير أنه ربما يحصل بالفعل على تقدير مختلف إلى حد ما عن التقدير المتنبأ به. ففي كل من هذين المثالين نلاحظ أن هناك أخطاء في القياس، و أخطاء في التنبؤ على الترتيب. وقد أوضحنا فيما سبق أن هناك مصادر متعددة لهذه الأخطاء بعضها أخطاء غير منتظمة Unsystematic Errors، والبعض الآخر أخطاء منتظمة Systematic Errors. فالأخطاء غير المنتظمة لا يكون لها نمطاً متسقاً، وتتذبذب من موقف اختبارى إلى آخر. فتشتت الانتباه، والتخمين، وانخفاض الدافعية تعد من مصادر الخطأ غير المنتظم. بينما الأخطاء المنتظمة يكون تأثيرها متكافىء على الفرد في مرتي تطبيق الاختبار، أو على جميع الأفراد الذين يطبق عليهم الاختبار في وقت معين. فالتدريب، والنسيان، والتعب، والنمو، يمكن اعتبارها مصادر خطأ منتظم، حيث إنها تعمل أحياناً على رفع الدرجات، وأحياناً أخرى تؤدي إلى انخفاضها. غير أن مصادر الخطأ المنتظم وغير المنتظم تكون متداخلة؛ فملاءمة الغرفة التى يطبق فيها الاختبار ربما تسهم في الأخطاء المنتظمة في بعض الأحيان إذا ما اختبر جميع الأفراد في غرفة سيئة التهوية أو الإضاءة مثلاً، أو تسهم في أحيان أخرى في الأخطاء غير المنتظمة إذا ما طبق الاختبار في غرفة متسعة ثم أعيد تطبيقه في غرفة ضيقة. ومفهوم ثبات درجات الاختبار Reliability الذي أوضحناه يتعلق بمشكلة ضبط وتقييم تأثير الأخطاء غير المنتظمة في الدرجات، بينما يتعلق مفهوم الصدق Validity بالإجراءات التجريبية المختلفة التى تستخدم في تحديد مدى تأثر الاختبار بالأخطاء المنتظمة. وهذا يتطلب توفر محك أو محكات لتقييم تواجد أو عدم تواجد هذه الأخطاء. لذلك فإنه ينبغي تحديد هذه المحكات، وإعدادها، وجمع بيانات حولها، لكي نتمكن من الحكم على الصدق. وتختلف طبيعة هذه المحكات التي تستخدم في تقييم صدق الاختبار باختلاف أغراض الاختبار الذي يتم بناؤه، واختلاف استخداماته.

أنواع صدق الاختبار:

حددت الجمعية الأمريكية لعلم النفس ثلاثة أغراض رئيسة للاختبارات واستخداماتها كالتالي:

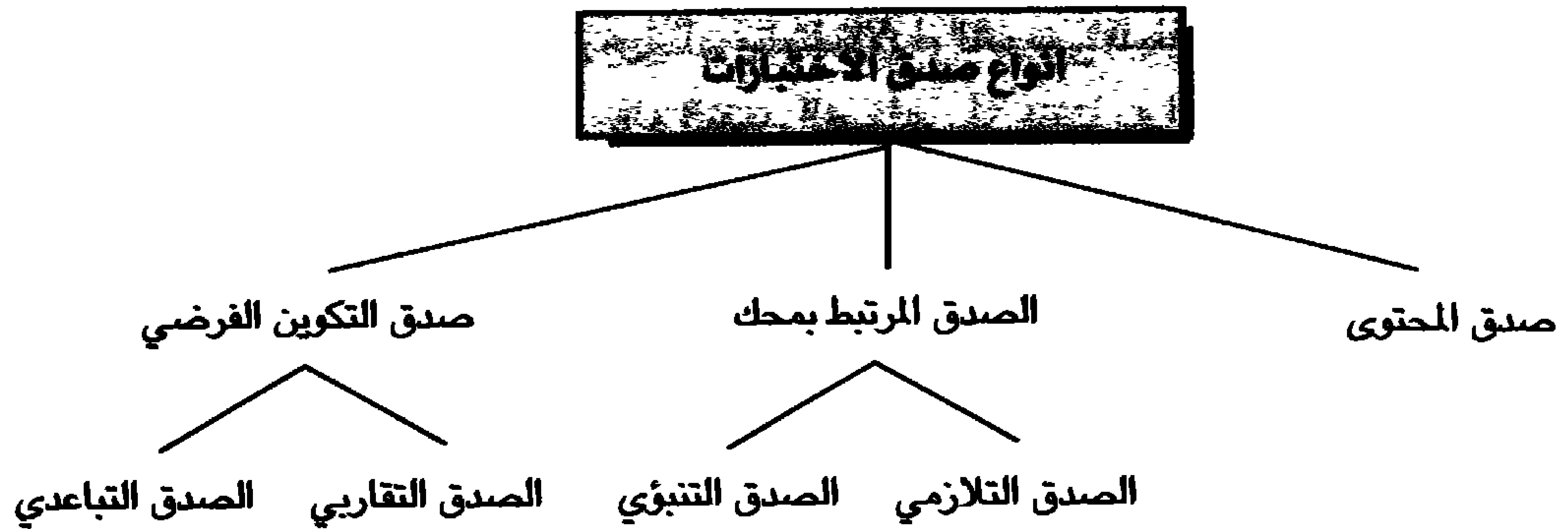
(1) تحديد كيفية أداء الفرد في نطاق شامل Universe من المهام أو المعارف أو المهارات التي يفترض أن الاختبار يشتمل على عينة ممثلة له. وهذا النوع من الصدق يسمى "صدق المحتوى Content Validity".

(2) التنبو بأداء مستقبلي، أو تقدير الوضع الراهن للفرد في تفسير معين يختلف عما يقيسه الاختبار. ومن أمثلة ذلك: التنبؤ بتقديرات الطلاب في الكلية من درجاتهم في اختبار الاستعداد الدراسي، أو التنبؤ بالحالة الانفعالية للفرد من بروفيل درجات الاختبارات الفرعية لاختبار ويكسلر لذكاء الراشدين. وهذا النوع من الصدق يسمي "الصدق المرتبط بمحك Criterion-related Validity".

(3) الاستدلال على درجة امتلاك الفرد لسمة أو خاصة أو تكوين فرضى ينعكس في الاختبار. فمثلاً ربما نود معرفة ما إذا كان الفرد يمتلك درجة مرتفعة من سمة مجردة معينة مثل الذكاء أو الابتكارية، ولا يمكن ملاحظة أى منهما ملاحظة مباشرة. ويمكن التوصل إلى معلومات تتعلق بهاتين السمتين لدى الفرد، أو تتعلق بدراسة الاختبار ذاته، أو علاقته بغيره من الاختبارات، أو ببناء نظرية سيكولوجية. فنحن هنا نهتم بما يسمى "صدق التكوين الفرضى "Construct Validity" للاختبار.

ونوع المعلومات المتعلقة بالصدق التي يتم الحصول عليها سوف تعتمد على الغرض من الاختبار وفيم تستخدم نتائجه؛ فالاختبار ربما يكون صادقاً لغرض أو أغراض معينة ولا يكون كذلك لأغراض أخرى. لذلك فإن تحديد أنواع التفسيرات لدرجات الاختبار تحديداً واضحاً، أو أنواع القرارات المراد اتخاذها استناداً إلى هذه الدرجات يعد أمراً أساسياً قبل التحقق من صدق الاختبار.

ويلخص الشكل التخطيطي (4 - 6) التالي هذه الأنواع الثلاثة من الصدق وفرعياتها:



شكل (4 - 6) يوضح أنواع صدق الاختبارات

ويعتمد بعض هذه الأنواع الثلاثة من الصدق على التحليل المنطقي للاختبار ومفرداته ومحتواها، والبعض الآخر يعتمد على أدلة إمبريقية وإحصائية. وسوف أوضح فيما يلي كل من هذه الأنواع بشيء من التفصيل:

صدق المحتوى Content Validity:

يتعلق صدق محتوى الاختبار بالحكم على مدى كفاية مفرداته كعينة ممثلة لنطاق محتوى أو أهداف يُفترض أن الاختبار يقيسها. وعلى الرغم من أن صدق المحتوى ربما ينطبق على جميع أنواع الاختبارات المعرفية ومقاييس الشخصية. إلا أنه يناسب بدرجة أكبر الاختبارات التحصيلية المقننة، أو التي يكتبها المعلم لطلابه. والحقيقة أنه لا يوجد مؤشر كمي أو عددي لصدق المحتوى، وإنما يتم تقييم هذا النوع من الصدق عن طريق فحص محتوى مفردات الاختبارات فحصاً منطقياً في ضوء ما يقيسه الاختبار التحصيلي من أهداف أو مستويات تعليمية Standards، وذلك لتعرف مدى كفاية مفرداته في قياس هذه الأهداف أو المستويات في ضوء محتوى مجال دراسي معين. وهذا يتطلب تحديد النطاق الشامل Universe للمفردات المراد معاينتها، أي يتم انتقاء عينة عشوائية أو عشوائية طبقية منها، وتضمينها في الاختبار. وفي حالة الاختبارات التحصيلية المقننة يحدد دليل الاختبار كيفية إجراء ذلك، أما في حالة الاختبارات الصفية التي يكتبها المعلم لطلابه، فإنه يستند إلى جدول مواصفات الاختبار الذي يقوم بإعداده كما سيتضح في الفصل الخامس.

الصدق المرتبط بمحك Criterion - related Validity:

تستخدم كثير من الاختبارات بخلاف الاختبارات التحصيلية في التنبؤ بسلوك مستقبلي

معين للأفراد. إذ ربما يود باحث تربوى أن يتنبأ بالأداء المدرسى للطلاب بمعلومية درجاتهم في اختبار للذكاء، أو يود باحث في علم النفس الصناعى أن يتنبأ بأداء الأفراد في عمل معين بمعلومية أدائهم في بطارية من اختبارات الاستعدادات، أو يود باحث في علم النفس الكلينيكي أن يتنبأ بقابلية علاج مرضى معينين باستخدام المعلومات التى يتم جمعها باستخدام أسلوب المقابلة الكلينيكية، أو مقياس مناسب للشخصية وهكذا. ففى مثل هذه الحالات تكون وظيفة الاختبار أو المقياس التنبؤ بالأداء في متغير يتضمن أداءً يختلف كيفياً عن الأداء الذى يتطلبه الاختبار، وهذا الأداء المتنبأ به يكون له أهمية عملية أو اجتماعية. والمتغير المتنبأ به باستخدام درجات الاختبار أو المقياس يسمى المحك Criterion. لذلك يُطلق على هذا النوع من صدق الاختبارات والمقاييس "الصدق المرتبط بمحك".

ونظراً لأن تقدير الصدق المرتبط بمحك يتضمن دائماً جمع بيانات إمبريقية عن العلاقة بين درجات الاختبار ودرجات المقياس المحك Criterion Measure، فإنه يشار أحياناً إلى هذا النوع من الصدق على أنه الصدق الإمبريقي Empirical Validity. والاهتمام الرئيس هنا ينصب على أداء الأفراد على المقياس المحك، حيث إن درجات الاختبار المراد التحقق من صدقها ليست هى موضع التركيز، وإنما تُستخدم في التنبؤ بسلوك محك ذى أهمية. لذلك فإن محتوى الاختبار لا يعد أمراً مهماً، كما أن مفرداته لا ينبغي بالضرورة أن تكون لها علاقة واضحة بالمقياس المحك، فما يهمنا هو أن يكون للاختبار صدق إمبريقي، أي أن درجاته تتنبأ بالفعل بالأداء في المحك.

ويمكن التمييز بين نوعين من هذا الصدق : الصدق التلازمي Concurrent Validity، والصدق التنبؤي Predictive Validity.

والفرق الرئيس بينهما يتعلق بالمدة الزمنية الفاصلة بين تطبيق الاختبار المراد التحقق من صدقه، وتطبيق المحك. فالصدق التلازمي يتطلب تطبيق كل من الاختبار والمقياس المحك في نفس الوقت تقريباً، والصدق التنبؤى يتطلب تطبيق المقياس المحك بعد انقضاء مدة زمنية مناسبة على تطبيق الاختبار. فاختبار الشخصية متعددة الأوجه (MMPI) مثلاً يفيد في تصنيف الاضطرابات العقلية، لأنه قد تبين أن الأفراد الذين يتم تشخيص حالة الاضطراب العقلى لديهم بواسطة الطبيب النفسى (المحك) يميلون إلى نمط معين من الدرجات في مجموعات معينة من مفردات أو فقرات هذا الاختبار. ونظراً لأن العلاقة بين الاختبار والمحك مرتفعة، فإن يمكن القول بأن اختبار الشخصية متعدد الأوجه يتميز بصدق تلازمي مرتفع.

أما اختبارات الذكاء والاستعدادات (الاختبارات المنبأة)، فإنها غالباً ترتبط ارتباطاً مرتفعاً بعلامات الطلاب، وتقديرات المعلمين، ودرجات الاختبارات التحصيلية. لذلك فإن يمكن تقدير الصدق التنبؤى لمثل هذه الاختبارات بإيجاد قيمة معامل الارتباط بين درجاتها (المحك)، ودرجات الطلبة أو علاماتهم في وقت لاحق. غير إن قيمة هذا المعامل لا تزيد في كثير من الأحيان عن 0.60، مما يعنى أن 36% من تباين درجات أو علامات الطلاب يمكن التنبؤ بها من درجات الاختبار المحك.

لذلك ينبغي الحيطة عند استخدام الاختبارات والمقاييس في التنبؤ بالسلوك المستقبلي. كما ينبغي مراعاة أن التنبؤ يتعلق بمجموعة من الأفراد وليس بأداء فرد معين. فمعامل الارتباط يعتمد على درجات اختبارين طبقا على مجموعة من الأفراد، وبالتالي يشير الصدق التنبؤي إلى صدق مجموعةٍ من الدرجات، وليس لدرجة فرد معين. وعلى الرغم من إمكانية التوصل إلى بعض الاستدلالات عن درجة الفرد، إلا أن طريقة جمع البيانات المتعلقة بالصدق،، وأخطاء القياس التي ربما تؤثر في كل درجة تقلل من الثقة في الاستدلالات الفردية، ولكنها يمكن أن تتميز بدرجة معقولة من الثقة إذا تعلقت بمجموعة الأفراد.

ويعتمد اختيار أي من نوعي الصدق المرتبط بمحك على ما إذا كنا نود التنبؤ بسلوك مستقبلى، أو تقييم الوضع الراهن للسلوك. وتستخدم العديد من المتغيرات كمقاييس محكة في تقييم الصدق المرتبط بمحك كما أشرنا في الأمثلة السابقة. واختيار المقياس المحك يعتمد على القرار المراد اتخاذه، ومجال السلوك المستقبلي موضع الاهتمام؛ فالمقياس المحك في المجال الصناعى أو مجالات الأعمال يختلف عنه في المجال التربوي والكلينيكي.

غير إنه في جميع الحالات ينبغي العناية باختيار المحك، وأن تتوافر فيه خصائص معينة. فالتحقق من الصدق المرتبط بمحك يتطلب أن تكون العلاقة بين المحك والاختبار كبيرة، لكي يتسنى اتخاذ قرار صائب، فإذا لم يعكس المقياس المحك السلوك أو الناتج المرجو، فإن القرار الذى يستند إلى هذا النوع من الصدق يكون أقل فاعلية

لذلك فإنه ينبغي التحقق من أن المقياس المحك يكون موائماً لطبيعة القرار المراد اتخـــاذه، والسلوك المراد التنبؤ به، وأن يؤدي إلى قياسات متسقة عبر الزمن واختـــلاف الظروف، ويكـــون خالياً بقدر الإمكان من عوامل التحيز والذاتية، بحيث لا تؤثر هذه العوامل في درجات الأفراد.

صدق التكوين الفرضي Construct Validity :

يتعلق صدق المحتوى، وصدق التكوين الفرضي باستخدامات عملية مفيدة لنتائج الاختبارات والمقاييس. فصدق المحتوى يفيد في تحديد مدى تمثيل درجات الاختبار لتحصيل الطلاب في مجال دراسي معين وما يتضمنه من نواتج تعليمية. والصدق المرتبط بمحك يفيد في تقدير أو التنبؤ بأداء حالى أو مستقبلي في مجال أكاديمي أو مهني معين، غير أنه في حالات أخرى ربما نود تفسير درجات الاختبارات في ضوء تكوينات فرضية Constructs، أو سمات نفسية Traits، أو خصائص عامة Qualities، مستمدة من نظرية سيكولوجية معينة، مثل: القلق، والاتجاه العلمي، والذكاء، والابتكارية، والدافعية، والتفكير الابداعي، وغير ذلك. فبدلاً من الاقتصار على معرفة درجة الفرد في اختبار حسابي معين، أو الاستناد إلى هذه الدرجة في التنبؤ بنجاحه في الرياضيات، ربما نود الاستدلال من درجته على أنه يمتلك درجة معينة من القدرة الاستدلالية مثلاً. وهذا يمكننا من وصف أداء الفرد وصفاً أكثر اتساعاً، ويساعد في العديد من الاستخدامات. فالتكوين الفرضى إذن هو صفة أو خاصية نفسية نفترض وجودها، وذلك لتفسير جوانب معينة من سلوك الأفراد. ونوع الصدق المناسب في هذه الحالة يسمى "صدق التكوين الفرضي".

ويتطلب تقدير هذا النوع من الصدق دراسات متعددة وأساليب منطقية، وإحصائية، وتجريبية.

وقد اقترح كرونباك ومهيل (Cronbach & Meehl, 1955) خمسة أنواع من الأدلة ينبغي جمعها لتأييد صدق التكوين الفرضي للاختبارات والمقاييس كالتالي:

(1) **الفروق بين مجموعات من الأفراد :** فعينات من الأفراد الذين يفترض اختلافهم في المتغير موضع البحث يمكن التنبؤ بتباين أدائهم، والبيانات التي يتم جمعها تكون مستعرضة بدرجة أساسية.

(2) **التغير في الأداء :** يمكن إجراء دراسات طولية لتعرف التغيرات عبر الزمن أو الناجمة عن اختلاف الظروف لنفس المجموعة من الأفراد.

(3) **الارتباط :** ينبغي أن ترتبط درجات اختبار معين بدرجات اختبار آخر يقيس المتغير نفسه، بينما لا ترتبط درجاته بدرجات اختبارات أخرى لا تقيس هذا المتغير. ففي الحالة

الأولى يمكن التوصل إلى الصدق التقاربي Convergent Valitdity، وفي الحالة الثانية يمكن التوصل إلى الصدق التمايزى Discriminant Validity للاختبار، وقد اقترح كامبل وفيسك (Campell & Fiske, 1959). هذين النوعين من الصدق، حيث يقدمان أدلة مهمة عن صدق التكوين الفرضى للاختبار. ولعل هذا يتطلب استخدام أسلوب التحليل العاملى Factor Analysis لمصفوفة معاملات الارتباط بين درجات الاختبار، ودرجات اختبارات ومتغيرات أخرى موائمة.

(4) الاتساق الداخلي: ويقصد به الارتباط بين درجات مفردات الاختبار، أى درجة قياس المفردات للسمة نفسها.

(5) دراسات عملية لطريقة تناول الفرد للاختبار: ويقصد بذلك ماذا يفعل الفرد عندما يطبق عليه الاختبار؟ وما العمليات العقلية التي تنطوي عليها استجابات الفرد لمفردات الاختبار؟ وهل تختلف استجابات الفرد لصيغ متكافئة من الاختبار؟ فهذه الأسئلة وغيرها ينبغي دراستها بأساليب تجريبية وإحصائية.

تعقيب عام على أنواع صدق الاختبارات والمقاييس:

من هذا يتضح أن أنواع الصدق الثلاثة ليست مستقلة؛ فالتحقق من صدق المحتوى ينبغي أن يسبق عادة الصدق المرتبط بمحك. والتوصل إلى أدلة عن صدق التكوين الفرضي يتطلب جمع بيانات مستمدة من تلك المستخدمة في التحقق من صدق المحتوى، والصدق المرتبط بمحك. كما أوضح كرونباك (Cronbach, 1971) ايضاً أنه يمكن التحقق من صدق المحتوى إمبريقياً بإيجاد الارتباط بين درجات صيغتين من صيغ الاختبار نفسه إذا استند بناؤهما إلى نفس نطاق المحتوى Content Domain .

وكذلك يمكن ضم اجراءات تقدير صدق التكوين الفرضي، وتقدير الثبات معاً للتوصل إلى أدلة عن صدق محتوى الاختبار.

وخلاصة القول أنه لا توجد طريقة واحدة للتحقق من صدق الاختبارات والمقاييس التربوية والنفسية، وذلك لأن نوع الأدلة المرجوة يعتمد على كيفية استخدام هذه الاختبارات والمقاييس؛ فالاختبار لا يكون صادقاً أو غير صادق بعامة، وإنما يمكن تحديد ذلك في ضوء أغراض بناء الاختبار واستخداماته. ومن المعلوم أن الاختبار ينبغي أن تتميز درجاته بالثبات قبل الحكم

على صدقه، والصدق ليس خاصة من خصائص الاختبار على وجه الخصوص، وإنما يتعلق بالاستدلال الذي يمكن التوصل إليه استناداً إلى درجات الاختبار. لذلك فإن الاهتمام ينبغي أن يُوجه بدرجة أساسية إلى صدق التفسير المستمد من هذه الدرجات.

ونظراً لأن البيانات المستمدة من دراسات الصدق تتعلق بمواقف معينة، فإن نتائج هذه الدراسات سوف تعتمد على خصائص هذه المواقف. ويترتب على ذلك أن يكون للاختبار أكثر من صدق واحد بقدر المواقف التي يتم دراسة الصدق في ضوئها. وتعميم نتائج صدق الاختبارات والمقاييس على عينات مختلفة من الأفراد مستمدة من مجتمع معين يتطلب إجراء دراسات مستعرضة للصدق Cross - Validation وذلك للتحقق من تماثل النتائج .

(ثالثاً) تحليل مفردات الاختبار وأهميته:

نظراً لأن نوعية الاختبار وجودته تعتمد اعتماداً كبيراً على نوعية وجودة المفردات التي يشتمل عليها، فإن تحليل المفردات Item Analysis يعد أمراً ضرورياً لتحسين الاختبارات، وبخاصة الاختبارات التحصيلية، سواء المقننة أو الصفية التي يكتبها المعلم لطلابه. وتحليل المفردات هو أسلوب منظم يصمم للحصول على معلومات محددة تتعلق بكل مفردة من مفردات الاختبار. وهذه المعلومات يمكن الاستفادة منها في تحديد المفردات الغامضة أو المربكة أو غير الفاعلة، من أجل مراجعتها أو استبعادها، وانتقاء افضل المفردات المتوافرة لتضمينها في الصيغة النهائية للاختبار. وكذلك يفيد تحليل إجابات الطلاب على كل مفردة من مفردات الاختبار في تعرف صعوبات التعلم لديهم لمعالجتها.

ويتضمن تحليل المفردات اساليب احصائية لتقدير مستوى صعوبة المفردة، ودرجة تمييزها بين الطلاب. ونظراً لأن معظم هذه الأساليب تقتصر على المفردات ثنائية الدرجة مثل: مفردات الاختيار من متعدد، والصواب أو الخطأ، والمزاوجة، فإنه يمكن ايضاً باستخدام اساليب تحليل المفردات تقييم فاعلية المشتتات Distractors، أو بدائل الاختيارات التي تكون جزءاً من بعض هذه المفردات.

وعلى الرغم من أن هناك أساليب إحصائية معقدة تستخدم في تحليل مفردات الاختبارات المقننة التي يقوم بإعدادها خبراء القياس والتقويم، وتقوم مؤسسات متخصصة بنشر هذه الاختبارات، إلا أنه توجد اساليب إحصائية بسيطة يمكن استخدامها في تحليل مفردات

الاختبارات التحصيلية الصفية التي يكتبها المعلم لطلابه، وهو ما سوف أوضحه في الجزء التالي. ويمكن للقارئ المهتم بالاساليب المتقدمة، وبخاصة التي تستند إلى نظرية الاستجابة للمفردة Item Response Theory (IRT) الرجوع إلى كتاب: القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة، وكتاب نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد، وتطبيقاتها في القياس النفسي والتربوي للمؤلف نفسه.

معامل صعوبة المفردة:

إن تحديد مستوى صعوبة كل مفردة من مفردات الاختبار يعد ضرورياً نظراً لأنه يبين للمعلم كيفية أداء الطالب في المهمة التي تقيسها المفردة، وكذلك المستوى العام لأداء طلبة صف معين في كل مفردة من مفردات الاختبار، وبذلك يستطيع المعلم تحديد مدى تحقق الأهداف التعليمية التي تقيسها هذه المفردات. كما أن معرفة مقدار معامل صعوبة المفردة يساعد في تعرف المفردات التي تكون غاية في الصعوبة أو السهولة بالنسبة لطلاب صف معين.

ويمكن التعبير عن صعوبة المفردة Item Difficulty بنسبة عدد الطلاب الذين أجابوا على المفردة إجابة صحيحة. ومقدار هذه النسبة يسمى "معامل الصعوبة Difficulty Index". فكلما زاد هذا المقدار دل ذلك على سهولة المفردة، وكلما قل اعتبرت المفردة صعبة.

اي أن معامل صعوبة المفردة = $\frac{\text{عدد الطلاب الذين أجابوا على المفردة إجابة صحيحة}}{\text{العدد الكلي للطلاب الذين طبق عليهم الاختبار}}$

(9 - 4) ....

فإذا كان عدد الطلاب الذين أجابوا على احدى المفردات إجابة صحيحة (24) مثلاً، والعدد الكلي للطلاب40، فإن معامل صعوبة هذه المفردة يساوي $\frac{24}{40} = 0.60$. أي أن هذه المفردة تعد متوسطة الصعوبة إلى حد ما بالنسبة لمجموعة الطلاب.ومقدار معامل الصعوبة يتراوح بين (صفر ، 1).

ومن الجدير بالذكر انه لا يوجد اتفاق بين خبراء القياس والتقويم التربوي على التوزيع الأمثل لمعاملات صعوبة اختبار معين، إذ أن هذا يعتمد جزئياً على نوع الاختبار المستخدم، أي

ما إذا كان الاختبار مرجعي المعيار NRT، أم مرجعي المحك CRT. فالاختبار من النوع الأول يهدف عادة للتمييز الأقصى بين المختبرين فيما يقيسه، لذلك ربما يكون من المناسب أن تتراوح صعوبة مفرداته بين (0.15 ، 0.85) أي أن تتوزع قيم معامل الصعوبة توزيعاً متعادلاً على مدى السمة أو القدرة المقاسة. ويرى بعض خبراء القياس أن الاختبار يمكن أن يميز تمييزاً أقصى بين الطلاب إذا كان متوسط مقدار معامل صعوبة مفرداته (0.50) . أما إذا كان الاختبار من النوع الثاني الذي لا يهتم بالتمييز بين الأفراد، وإنما تحديد مدى تحقيق كل طالب للأهداف التعليمية المتعلقة بمجال دراسي معين، ففي هذه الحالة تقيس مفردات الاختبار هذه الأهداف قياساً مباشراً، وتتحدد صعوبة المفردة وفقاً للمستوى المعرفي للهدف.

فإذا كان الهدف التعليمي من مستوى التذكر مثلاً، فإن المفردة التي تقيس هذا الهدف تكون بالطبع أسهل من المفردة التي تقيس هدفاً تعليمياً من مستوى التحليل أو التقويم. وفي هذه الحالة لا يكون ضرورياً الالتزام بأن يكون متوسط صعوبة المفردات (0.50) كما في الحالة الأولى.

ويمكن القول بعامة أن مستوى صعوبة مفردة اختبار يعتمد على الهدف من استخدام نتائج الاختبار، وطبيعة مجموعة الأفراد التي يطبق عليها الاختبار. فمقدار معامل الصعوبة يعتمد على هذه المجموعة، ويؤدي تغيرها إلى تغير هذا المقدار نظراً لتغير قدرة أفراد المجموعة المراد اختبارها.

معامل تمييز المفردة:

إن المفردات الاختبارية التي تتساوى في مقدار صعوبتها لا تميز بالضرورة بنفس القدر بين المستويات التحصيلية المختلفة للطلاب فيما يقيسه الاختبار. لذلك ينبغي إيجاد درجة تمييز كل مفردة للتحقق من فاعليتها في التمييز بين الطلاب أقوياء التحصيل، والطلاب ضعاف التحصيل.

فإذا وجد المعلم مثلاً أن نسبة الطلاب الذين أجابوا على مفردة معينة إجابة صحيحة في مجموعة الطلاب من ذوي الدرجات الكلية المنخفضة في الاختبار تساوي نسبة من أجابوا على المفردة نفسها إجابة صحيحة في مجموعة الطلاب من ذوي الدرجات الكلية المرتفعة، فإن مثل هذه المفردة لا تكون قد ميزت بين المجموعتين الطرفيتين، وتحتاج بلا شك إلى مراجعة.

ويمكن الحصول على مقدار معامل تمييز المفردة Discrimination Index بإيجاد الفرق بين نسبة عدد طلاب المجموعة العليا (ذوي الدرجات الكلية المرتفعة في الاختبار) الذين أجابوا على المفردة إجابة صحيحة، ونسبة عدد طلاب المجموعة الدنيا (ذوي الدرجات الكلية المنخفضة في الاختبار) الذين أجابوا على المفردة إجابة صحيحة.

ولكن كيف يمكن تحديد المجموعتين الطرفيتين العليا والدنيا؟ لقد أوضح كيلي Kelley أن اختيار نسبة عدد الأفراد في كل من هاتين المجموعتين بحيث تساوي (10%) مثلاً ربما يقلل من ثبات الدرجات، وذلك لأن عدد أفراد كل من المجموعتين سوف يكون قليلاً. وقد وجد انه في التوزيع الاعتدالي تكون النسبة الفاصلة المثلى لكل من المجموعتين هي (27%). أما إذا كان عدد طلاب الصف قليلاً، فإن أخطاء المعاينات في معامل التمييز سوف تكون كبيرة نسبياً. لذلك ينبغي الحيطة عند اختيار نسبة عدد أفراد المجموعتين الطرفيتين. وترى أنستازي Anastasi أن أي نسبة تتراوح بين (25% , 33%) سوف تكون مرضية.

ولايجاد مقدار معامل تمييز مفردة ينبغي أولاً ترتيب أوراق الإجابة بحسب الدرجات الكلية ترتيباً تنازلياً، ثم نفصل (27%) من الأوراق التي حازت على أعلى الدرجات (المجموعة العليا)، (27%) التي حازت على أقل الدرجات (المجموعة الدنيا) ثم نتبع الخطوات التالية:

(1) إيجاد عدد طلاب كل من المجموعتين العليا والدنيا الذين أجابوا على كل بديل من بدائل مفردة الاختيار من متعدد. ويمكن تكوين جدول تحليل المفردة كالتالي مثلاً:

الاجابة الصحيحة	رقم المفردة	المجموع	البدائل				
			أ	ب	جـ*	د	هـ
جـ	1	العليا	1	-	8	-	1
		الدنيا	1	-	6	2	1

جدول (4 - 1): يوضح نتائج تحليل مفردة اختبارية من نوع الاختيار من متعدد

ويلاحظ من الجدول أن:

- عدد طلاب المجموعة العليا (27%) = 10.
- وعدد طلاب المجموعة الدنيا (27%) = 10.
- وعدد طلاب المجموعة العليا الذين أجابوا على المفردة اجابة صحيحة = 8.
- وعدد طلاب المجموعة الدنيا الذين أجابوا على المفردة اجابة صحيحة = 6 .

(2) نوجد الفرق بين نسبة عدد طلاب المجموعة العليا الذين أجابوا على المفردة إجابة صحيحة، وعدد طلاب المجموعة الدنيا الذين أجابوا على المفردة إجابة صحيحة، للحصول على مقدار معامل التمييز.

ويتضح من الجدول أن:

$$\text{معامل التمييز} = \frac{8}{10} - \frac{6}{10} = \frac{2}{10} = +0.20$$

وعلى الرغم من أن المفردة تميز تمييزاً موجباً (+0.20) إلا أن مقدار معامل تمييزها منخفضاً، مما يستدعي مراجعة البدائل المختلفة للاجابات في المفردة.

تفسير معامل التمييز:

تقدم قيمة معامل التمييز للمعلم مؤشرات عن فاعلية كل مفردة في الاختبار، وما تتضمنه من بدائل أو اختيارات، وتتراوح هذه القيم بين (-1) ، (+ 1) .

فإذا كانت قيمة هذا المعامل (0.40) أو أكبر، فإن هذا يكون دليلاً على أن المفردة تميز بين المجموعتين بدرجة جيدة.

أما إذا كانت هذه القيمة تتراوح بين (0.20 ، 0.40)، فإن التمييز يكون لا بأس به. وإذا قلت عن (0.20) يكون التمييز ضعيفاً.

وإذا كانت قيمة التمييز صفراً، فإن هذا يعني أن المفردة غاية في السهولة أو الصعوبة، ولا تميز بين مرتفعي ومنخفضي التحصيل. فإذا كان الاختبار مرجعي المعيار (NRT) ويهدف للتمييز بين الطلاب فيما يقيسه، فإن هذه المفردة يجب استبعادها من الاختبار. أما إذا كان

الاختبار مرجعي المحك (CRT) ويقيس تمكن الطلاب من مهارة أساسية، فإن المفردة السهلة التي تمييزها صفراً لا ينبغي استبعادها، لأنها تكون مناسبة للمستوى المعرفي للهدف الذي تقيسه، واستطاع جميع الطلاب الإجابة عليها.

ومما لا شك فيه أن قيمة معامل التمييز تعتمد على درجة صعوبة المفردة الاختبارية، فكلما ابتعدت قيمة معامل الصعوبة عن (0.50) تقل القيمة القصوى لمعامل التمييز. لذلك فإنه إذا استخدم المعلم مفردات غاية في السهولة أو الصعوبة بالنسبة لمجموعة معينة من الطلاب، فإنه لا يجب أن يتوقع الحصول على قيم مرتفعة لمعاملات التمييز.

ولكن يجب أن نؤكد أنه إذا كان الهدف من الاختبار قياس إتقان الطلاب لمحتوى دراسي أو مهارة معينة، فإن المعلم ربما يجد أن مقدار معامل الصعوبة (0.80) مثلاً، ومعامل التمييز صفراً، ولا مانع مّن ذلك ما دام المعلم ليس مهتماً بإبراز الفروق الفردية بين الطلاب، وإنما يهدف لقياس درجة تمكنهم من المحتوى أو المهارة المرجوة.

الباب الثاني 2

أدوات القياس التربوي والنفسي

- الفصل الخامس: **الاختبارات التحصيلية**
- الفصل السادس: **اختبارات الذكاء العام**
- الفصل السابع : **اختبارات الاستعدادات الخاصة**
- الفصل الثامن: **اختبارات ومقاييس الشخصية**
- الفصل التاسع: **اختيار وتقييم الاختبارات والمقاييس**

الفصل الخامس

الاختبارات التحصيلية

- مقدمة
- مفهوم الاختبار التحصيلي
- الفروض الأساسية التي يرتكن إليها قياس التحصيل
- تصنيف الاختبارات التحصيلية
- استخدامات الاختبارات التحصيلية
- مراحل بناء الاختبارات الصفية
- الاختبارات التحصيلية المقننة
- خصائص الاختبارات التحصيلية المقننة
- بناء الاختبارات التحصيلية المقننة
- الفروق بين الاختبارات التحصيلية المقننة، والاختبارات الصفية

مقدمة:

عرضنا في الباب الأول أهم الأسس المنهجية والإحصائية التي يستند إليها القياس النفسي والتربوي بعامة، والاختبارات والمقاييس بخاصة، وكذلك أخلاقيات استخدام الاختبارات. وسوف أتناول في هذا الباب أدوات القياس التربوي والنفسي المتعلقة بالتحصيل الدراسي، والذكاء العام، والاستعدادات الخاصة، والشخصية الانسانية. وعلى الرغم من أن هذه الجوانب المعرفية وغير المعرفية أو الوجدانية تعد جوانب متكاملة للسلوك الإنساني، إلا إنني سأتناول كل منها في فصل مستقل لمزيد من التوضيح لهذه الجوانب أو السمات الانسانية الرئيسة، وما يتعلق بها من اختبارات ومقاييس متنوعة في ضوء المفاهيم والمبادىء التي تناولتها بالتفصيل في الفصول الأربعة السابقة. وينتهي هذا الباب بفصل مستقل عن كيفية اختيار وتقييم الاختبارات والمقاييس التربوية والنفسية. وسوف أستهل هذا الباب بالاختبارات التحصيلية لأنها أكثر شيوعاً واستخداماً في المؤسسات التربوية والتدريبية. ويتناول هذا الفصل مفهوم الاختبار التحصيلي، والفروض الأساسية التي يستند إليها قياس التحصيل، وتصنيف الاختبارات التحصيلية واستخداماتها المتعددة، ومراحل بناء الاختبارات التحصيلية الصفية والاختبارات التحصيلية المقننة، والفروق بين نوعي الاختبارات.

مفهوم الاختبار التحصيلي Achievement Test :

يُستخدم مفهوم التحصيل Achievement للإشارة إلى درجة أو مستوى النجاح الذي يحرزه الطالب في مجال دراسي عام أو متخصص؛ فهو يمثل اكتساب المعارف والمهارات والقدرة على استخدامها في مواقف حالية أو مستقبلية. ويعد التحصيل هو الناتج النهائي للتعلم. ويتأثر مستوى التحصيل والأداء بعوامل متعددة توجد وقت التعلم، كما يكون لها تأثير وسيط ما بين التعلم واستخدامات نواتجه، ويستخدم المعلمون مفهوم التحصيل للإشارة إلى قدرة الطلاب على تحقيق الأهداف التعليمية للمواد الدراسية، على الرغم من أن هذا يعد

استخداماً محدوداً للمفهوم، حيث إنه يمكن أن ينطبق على مختلف مجالات السلوك الانساني الفاعل. (Bertrand & Cebula, 1980, Ward & Ward, 1999). وهناك أدلة متعددة مستمدة من البحوث النفسية والتربوية تؤكد العلاقة الموجبة بين الذكاء والتحصيل. كما أن الدافعية، ومفهوم الذات، والضبط الذاتي، ومستوى طموح الطالب تؤثر في مستوى وجودة تحصيله. وكذلك تؤثر البيئة المنزلية والمدرسية، واتجاهات الآباء نحو إمكانات أبنائهم، وتوقعات المعلمين عن طلابهم في مستوى التحصيل لديهم. وقد بينت دراسة جوخ (Gough, 1953) وتيهان (Tea-han, 1963)وإينز (Innes, 1972)، وجنسن، وفيورستاين (Jensen & Feuerstein, 1986) أن السمات الشخصية المختلفة تؤثر تأثيراً واضحاً في التحصيل الدراسي للطلاب في مختلف مراحل التعليم. كما أن طلاب المدارس الثانوية مرتفعي التحصيل يتسمون بالكفاءة العقلية، والضبط الذاتي، والثراء المعرفي، والاستقلالية، والاتساق، والمثابرة. ويؤثر المناخ النفسي الذي يوفره المعلمون في سلوك الطلبة، الذي يؤثر بدوره في دافعيتهم، ومفهوم الذات لديهم، واتجاههم نحو المدرسة، فإذا كان هذا المناخ ايجابياً وبناءً، فإنه يساعد في تيسير تعلم الطلبة ورفع مستوى تحصيلهم، ولكنه يعوق ويحد من التحصيل إذا كان سلبياً.

وربما يتساءل البعض عن كيفية اختلاف الاختبارات التحصيلية عن غيرها من الاختبارات النفسية والتربوية إذ ربما يعتقد البعض أن الاختبارات التحصيلية تقيس ما تعلمه الطالب في المواد الدراسية. وعلى الرغم من أن هذا يبدو معقولاً، إلا أن معظم الاختبارات التربوية والنفسية تقيس أيضاً ما تعلمه الفرد، فاختبار قيادة السيارة مثلاً، يعد اختباراً تحصيلياً على الرغم من أنه لا يرتبط بمادة دراسية معينة. وربما يرى البعض أن الاختبارات التحصيلية تتميز بأنها تتطلب عمليات معرفية معينة للإجابة على مفرداتها، مثل: تذكر الحقائق، والاستدلال، وحل المشكلات وغير ذلك، غير أنه إذا فحصنا اختبارين من الاختبارات التي تقيس الجوانب العقلية أو المعرفية احدهما يقيس التحصيل والآخر يقيس الاستعداد الدراسي ربما يصعب التمييز بينهما تمييزاً واضحاً.

لذلك يرى براون (Brown, 1970) أن الاختبار التحصيلي يقيس التعلم الذي يحدث نتيجة للخبرة في موقف تعلم يتم التحكم فيه، مثل: صف مدرسي، أو برنامج تدريبي، وأن الإطار المرجعي الذي نهتم به هو الحاضر أو الماضي، أي الوضع الراهن لما تعلمه الطلاب.

الفروض الأساسية التي يرتكن اليها قياس التحصيل:

في ضوء تعريف مفهوم التحصيل الذي أوضحناه يتبين أن الاختبارات التحصيلية تصمم

لقياس المعارف والمهارات التي اكتسبها الطلاب في مجال محدود نسبياً. وهذا المجال قد يكون محدوداً جداً، مثل: تحصيل الطلاب لموضوع دراسي يتعلق بجغرافية الدول العربية، أو احدى العمليات الحسابية الأربع، أو يكون عريضاً ومتسعاً بحيث يشمل كل ما يدرس خلال العام الدراسي، مثل: رياضيات المدرسة الثانوية. إلا أننا في الحالتين نحاول قياس ما يعرفه الطالب، أو ما يمكنه اداؤه في نهاية مدة دراسية معينة. ونهتم بما تعلمه نتيجة لسلسلة معينة من الخبرات التي اكتسبها من دراسة منهج دراسي معين. ونظراً لأنه لا يمكن أن نقيس جميع المهارات الممكنة، فإن مفردات الاختبار تكون بمثابة عينة ممثلة لجميع المفردات المتعلقة بالمجال الدراسي المعين.

فإذا اخذنا بهذه النظرة عن مفهوم التحصيل، فإننا في الحقيقة نفترض بعض الافتراضات المتعلقة بطريقة القياس كالتالي (Anastasi, 1976, Brown, 1970; Elliot, 1984):

(1) يمكن تحديد محتوى أو نطاق المعارف والمهارات التي يشتمل عليها الاختبار في عبارات سلوكية. وعلى الرغم من أن بعض المربين يرون أن هناك بعض الأهداف التربوية ربما يصعب تحديدها بطريقة سلوكية ، مثل: الاتجاه الايجابي نحو التعلم، أو التذوق الفني، إلا أن الأهداف التربوية ينبغي أن تحدد التوقعات أو المستويات Standards المرجوة للمادة الدراسية لكي نتمكن من قياسها.

(2) يقيس الاختبار الأهداف أو التوقعات أو المستويات المهمة، ولا يقيس اهدافاً غير ضرورية. وهذا الافتراض يتعلق بصدق محتوى الاختبار Content Validity الذي اوضحناه في الفصل السابق. فمثلاً لكي يحل الطالب مسألة جبرية، فإنه يحتاج إلى عدد من المهارات، مثل: قراءة المسألة وفهمها، وترجمتها إلى صيغة جبرية، والقيام بالعمليات الحسابية المطلوبة. فإذا كان الاختبار قد صُمم لقياس قدرة الطالب على حل المسائل الجبرية الكلامية، فعندئذ يكون الاختبار قد صُمم لقياس الغرض المطلوب.

أما إذا كنا نهتم فقط بقدرة الطالب على استخدام الصيغ الجبرية، فإن الاختبار السابق يكون مقياساً لقدرة الطالب على القراءة، ويصبح الاختبار لا يقيس الغرض الذي صُمم من أجله.

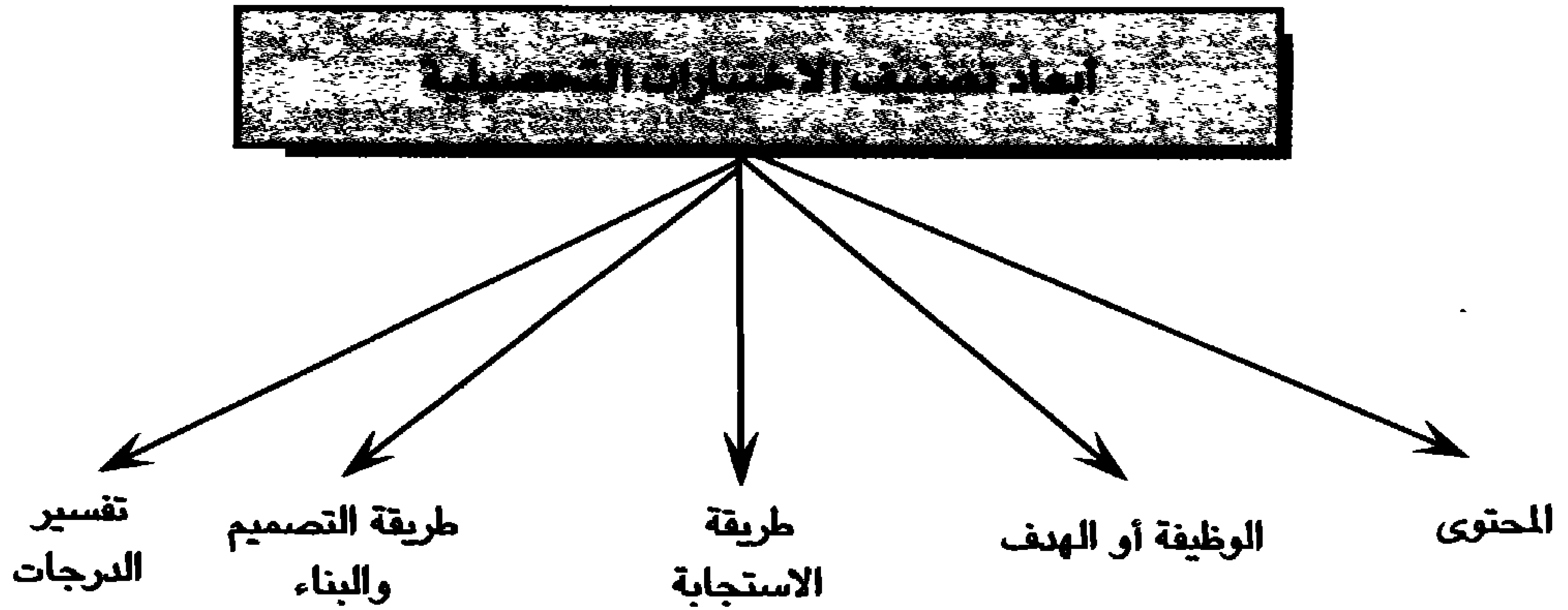
(3) توفر فرصة كافية للطلاب الذين سيطبق عليهم الاختبار لتعلم المادة التي سوف يشتمل

عليها الاختبار ويقيسها. وبتعبير آخر، إذا أردنا المقارنة بين الطلاب أو التحقق من اتقانهم للأهداف أو المستويات التعليمية المرجوة، فإنه يجب أن توفر لهم الفرص المتكافئة للتعلم. وهذه الفرص تتحقق بدرجة معقولة في الاختبارات الصفية التي يكتبها المعلم لطلابه، لأن جميع طلاب صف مدرس معين يدرسون المنهج نفسه، ويحدد لهم المعلم محتوى وأهداف الاختبار.

إلا أن عدم تكافؤ الخبرات التعليمية تعد مشكلة في حالة الاختبارات التحصيلية المقننة التي سأوضحها في هذا الفصل، حيث يقيس الاختبار محتوى تعليمي ربما يختلف من مدرسة إلى اخرى، وتتباين في كل منها خبرات الطلاب، أي تكون خبراتهم غير متكافئة مع خبرات طلاب المجموعة المرجعية Reference Group التي تم تقنين الاختبار عليها.

تصنيف الاختبارات التحصيلية:

يمكن تصنيف الاختبارات التحصيلية وفقاً لعدد كبير من الأبعاد، لعل أهمها ما يوضحه الشكل التخطيطي (5 - 1) التالي:



شكل (5 - 1) يوضح الأبعاد المختلفة لتصنيف الاختبارات التحصيلية

وفيما يلي توضيحاً لكل من هذه الابعاد:

المحتوى Content: تختلف الاختبارات التحصيلية في محتواها؛ فبعض الاختبارات - كما سبق أن ذكرنا - يكون محدود المحتوى؛ مثل: قياس حالة الطقس، أو تكون متسعة وعريضة في محتواها، مثل: مهارات القراءة، أو حل المشكلات. وبعض الاختبارات التحصيلية

يكون على شكل بطاريات Batteries of Tests تقيس مواداً مختلفة بدلاً من مادة واحدة. مثل اختبار أيوا للمهارات الاساسية (ITBS)، واختبارات كاليفورنيا التحصيلية (CAT)، والاختبارات المتتابعة للتقدم التربوي (STEP) وغيرها. فهذه الاختبارات تركز على مهارات تربوية مثل: القراءة، والحساب، واستخدام اللغة، ومهارات الاستذكار المتعلقة بقراءة الخرائط والرسوم التوضيحية والأشكال البيانية، واستخدام المصادر المتنوعة، وغير ذلك.

وهنا يحسن تصنيف الاختبار على أساس الفرقة الدراسية، أي بطارية اختبار تحصيلي للمرحلة الأولى، أو المرحلة الإعدادية، أو الثانوية.

وظيفة أو هدف الاختبار Purpose: يمكن تصنيف الاختبارات وفقاً للأغراض التي يُستخدم فيها وهي: الاختبارات المسحية Survey Tests، والاختبارات التشخيصية Diag-nostic Tests واختبارات الاستعداد أو المقدرة Readiness Tests. فالاختبار التحصيلي يعد اختباراً مسحياً إذا استُخدم في تقدير المستوى العام لأداء الفرد في مادة دراسية أو محتوى دراسي معين. لذلك يمكن استخدامه في إعطاء صورة تقريبية عن مكانة الفرد بالنسبة لأقرانه فيما يقيسه الاختبار، أي مقارنة أداء الطالب بأداء اقرانه. ويمكن أن يكون الاختبار المسحي على شكل بطارية من الاختبارات المسحية الفرعية التي يتعلق كل منها بمادة دراسية مختلفة تناسب فرقة معينة Achievement Batteries. ومن أمثلة ذلك اختبار ستانفورد التحصيلي Stanford Achievement Test, (SAT) الذي أجريت عليه تغييرات كثيرة في طبعاته المتتالية. وهذا الاختبار يناسب صفوف المرحلتين الابتدائية والمتوسطة، ويشتمل على (6) بطاريات يحتاج تطبيق كل منها حوالي (5) ساعات يمكن تقسيمها إلى جلسات متعددة، وتتكون الاختبارات الفرعية للبطارية من صيغتين متكافئتين، ويوجد دليل إرشادي للتدرب على الاختبارات التي تشتمل عليها قبل التطبيق الفعلي لأي منها. فمثلاً بطارية المرحلة المتوسطة تشتمل على (11) اختباراً فرعياً تتعلق بالمفردات اللغوية، وفهم المقروء، ومهارات التعامل مع الكلمات، ومفاهيم رياضية، وتطبيقات رياضية، وتهجي كلمات، واستخدام القواعد اللغوية، والمهارات المستخدمة في العلوم الاجتماعية، مثل (مهارات الاستدلال المتعلقة بالجغرافيا، والتاريخ والاقتصاد، والعلوم السياسية، وعلم الاجناس البشرية، وعلم الاجتماع)، ومفاهيم علمية، وفهم الموضوعات التي تقدم شفوياً. وجميع مفردات هذه الاختبارات من نوع الاختيار من متعدد. كما يمكن تفسير الدرجات استناداً إلى الأهداف التعليمية، التي تقيسها كل مفردة

أو مجموعة من المفردات، وكذلك يمكن تفسيرها استناداً إلى معايير مئينية وتساعيات معيارية تم اعدادها لكل اختبار، وقد سبق توضيح هذه المعايير في الفصل الثالث وبذلك تسمح هذه البطاريات من الاختبارات بالتفسير مرجعي المحك (CRT)، والتفسير مرجعي المعيار (NRT). وقد تم تقنين هذه البطاريات فيما بعد على عينة قومية في مدارس الولايات المتحدة الأمريكية قوامها 275,000 طالب ممثلة لكل شرائح المجتمع، وأصبح لهذه البطاريات معايير قومية National Norms.

ونظراً لان الاختبار المسحي يكون محتواه متسعاً، فإنه يصعب باستخدامه إعطاء صورة صادقة عن جوانب القوة والضعف في تحصيل الطلاب.

وعلى العكس فإن الاختبار التشخيصي يقيس التحصيل بدرجة أكثر تفصيلاً وعمقاً نظراً لتركيزه على قياس مكونات المادة الدراسية، ويُستخدم في تحديد جوانب القوة والضعف في تحصيل الطلاب. وينقسم عادة إلى عدد من الاختبارات الفرعية، وتقدر درجة لكل طالب في كل اختبار فرعي. ولذلك فهو يكون اطول من الاختبار المسحي. فمثلاً الاختبار التشخيصي في القراءة ربما يشتمل على اختبارات فرعية احدها يقيس تعرف الكلمات، وآخر يقيس فهم معاني الكلمات، وثالث يقيس سرعة القراءة، ورابع يقيس فهم الموضوع المقروء، وغيرها. وهذا يتطلب تحليل مكونات مهارة القراءة إلى عناصرها الرئيسة والفرعية التي يتم قياسها. ويُستخدم الاختبار التشخيصي عادة بعد ما يتضح من تطبيق الاختبار المسحي ضعف مستوى أداء الطلاب، وبذلك نستطيع عن طريق استخدام الاختبار التشخيصي التحديد الدقيق لجوانب الضعف، وتقديم النوع المناسب من التعليم العلاجي.

وتوجد العديد من الاختبارات التشخيصية المنشورة باللغة الانجليزية مثل: اختبار ستانفورد التشخيصي في القراءة (SDRT)، وكذلك في الرياضيات (SDMT)، واختبار كاليفورنيا التشخيصي في الحساب (CDAT)، وغيرها. وهذه الاختبارات تناسب الطلاب في مختلف مراحل التعليم العام. وقد أعد مؤلف هذا الكتاب مجموعة من الاختبارات التشخيصية التي تقيس مهارات أساسية مختلفة لدى طلاب بعض الكليات (علام، 2002 , 2004).

أما اختبار المقدرة فهو يبين ما إذا كان الطالب لديه المهارة المطلوبة لتعلم المادة التعليمية اللاحقة. وبتركيز اختبار المقدرة على المهارات المهمة اللازمة للتعلم اللاحق، فإن اختبار المقدرة يشبه في ذلك الاختبار التشخيصي، وباستخدامه للتنبؤ بالأداء اللاحق، فإنه يشبه اختبار الاستعداد الدراسي.

وأفضل مثال لهذا النوع من الاختبارات هو اختبار الاستعداد للقراءة Reading Readiness الذي يطبق على أطفال الحضانة الذين يزمعون الالتحاق بالصف الأول الابتدائي. ويستخدم هذا الاختبار في تحديد ما إذا كان الطفل قد تعلم المهارات الضرورية للبدء في تعلم القراءة.

طريقة الاستجابة Response Type :

يمكن أيضاً تصنيف الاختبارات التحصيلية وفقاً لطريقة الاستجابة لمفرداتها أو المهام التي تشتمل عليها. فعلى الرغم من أن كثيراً من الاختبارات التحصيلية تعد من نوع الورقة والقلم Paper - and- Pencil Test، إلا انه توجد اختبارات تتطلب مهارات ادائية مختلفة Perfor- mance Tests، مثل: اختبارات الأداء الموسيقي، والاختبارات العملية، واختبارات مهارات تشغيل اجهزة معينة، وغير ذلك. ولعل هذا النوع من الاختبارات يناسب تقويم الأداء Perfor- mance Assessments الذي أصبح ينادي به خبراء القياس والتقويم في وقتنا الحاضر، في إطار ما يسمى بالتقويم البديل أو التقويم الواقعي Authentic Assessments. ويمكن للقارئ الرجوع إلى كتاب "التقويم التربوي البديل: أسسه النظرية والمنهجية وتطبيقاته الميدانية للمؤلف، وذلك للاطلاع على تفاصيل هذا المجال المستحدث من مجالات القياس والتقويم التربوي المعاصر.

وتختلف اختبارات الورقة والقلم باختلاف انواع مفرداتها. إذ يمكن أن تشتمل على مفردات تتطلب التعرف من بين بدائل اجابات معطاة، مثل: مفردات الاختيار من متعدد، والصواب أو الخطأ، والمزاوجة. وكذلك يمكن أن تشتمل على مفردات أو أسئلة تتطلب إجابات مفتوحة دون تقديم بدائل اجابات معينة، حيث يجيب الطالب بما يراه مناسباً للسؤال، مثل أسئلة المقال بأنواعها المختلفة.

طريقة تصميم وبناء الاختبار Test Design :

يمكن أيضاً تصنيف الاختبارات التحصيلية إلى نوعين رئيسين: الاختبارات المقننة Standardized Tests، والاختبارات الصفية التي يكتبها المعلم Teacher - Made Tests. فالاختبار التحصيلي المقنن يعد بواسطة خبراء القياس ومؤسسات نشر الاختبارات، ويُستخدم في مختلف المدارس. ويختلف عن الاختبار الذي يكتبه المعلم في أن محتواه يكون متسعاً

بحيث يتفق مع الأهداف العريضة للمادة الدراسية. وبذلك يقيس معلومات اتفق خبراء التربية على أنها ضرورية لمتوسط الأداء في مادة دراسية معينة أو فرقة معينة. ويُطبق الاختبار المقنن على عينة كبيرة نسبياً من الطلاب خصائصها معلومة ويتم انتقاؤها بطريقة عشوائية، وذلك لايجاد معايير الاختبار استناداً إلى متوسط أداء هذه العينة، وتصبح هذه المعايير الوسيلة التي يقارن في ضوئها أداء كل طالب بمتوسط أداء هذه العينة. أما الاختبارات التي يكتبها المعلم لطلابه فلا تتطلب هذه المعايير، نظراً لأنها تتعلق بالطلاب الذين يقوم المعلم بالتدريس لهم، وتقيس الأهداف التعليمية المباشرة لما تم تعلمه. لذلك فإن محتواها يكون أكثر تحديداً. وسوف اتناول كل من هذين النوعين الرئيسيين من الاختبارات التحصيلية بعد قليل.

تفسير درجات الاختبار Score Interpretation :

يمكن تصنيف الاختبارات التحصيلية في ضوء مرجعية تفسير درجاتها. فإذا كان الاهتمام منصباً على تمييز الفروق الفردية بين الطلاب في مجال دراسي معين استناداً إلى درجات الاختبار، فإن الاختبار يسمى عندئذ "الاختبار مرجعي المعيار Norm - Referenced Test (NRT)، اما إذا انصب الاهتمام على معرفة مدى اتقان الطالب لمهارات معينة محددة في إطار محتوى دراسي معين، فإننا نقارن درجة الفرد بمحك أو مستوى أداء سبق تحديده بدلاً من مقارنته بأقرانه. ويسمى الاختبار في هذه الحالة "الاختبار مرجعي المحك Criterion - Referenced Test (CRT). وقد سبق أن اشرنا إلى هذين النوعين من الاختبارات في بعض المواضع في الباب الأول.

فالاختبار مرجعي المعيار هو ذلك الاختبار الذي يستخدم في تقييم أداء الفرد بالنسبة لأداء الافراد الآخرين في القدرة أو السمة التي يقيسها الاختبار. والهدف من ذلك ترتيب الأفراد بعضهم بالنسبة للبعض الآخر من حيث درجاتهم في الاختبار. ويكون للدرجة تفسيراً فقط في ضوء معيار الجماعة Test Norm أي متوسط أداء الجماعة في الاختبار.

وتصبح الدرجة لا معنى لها ويصعب تفسيرها إذا أغفلنا رد هذه الدرجة إلى هذا المعيار. وتعد الاختبارات المقننة اوالمسحية التي سبق ذكرها من هذا النوع.

أما الاختبار مرجعي المحك فيعرَّف بأنه ذلك الاختبار الذي يُستخدم في تقييم أداء الفرد بالنسبة إلى محك أو مستوى أداء مطلق دون الحاجة إلى مقارنة أدائه بأداء الأفراد الآخرين.

ويمدنا هذا النوع من الاختبارات بمعلومات محددة وتفصيلية عن تحصيل كل طالب في موضوع دراسي معين أو مهارة معينة. ودرجة الفرد في الاختبار مرجعي المحك تعبر عن مدى تحقيقه للأهداف التعليمية المرجوة. لذلك فإن هذا النوع من الاختبارات له قيمة تشخيصية لتحديد جوانب القوة والضعف في تحصيل كل طالب لمجموعة محددة من المعارف والمهارات، وبذلك يكون أقرب إلى الاختبارات التي يكتبها المعلم لطلابه.

استخدامات الاختبارات التحصيلية:

تُستخدم الاختبارات التحصيلية استخدامات متعددة لعل أهمها أنها تحدد ما يعرفه الطالب وما اكتسبه من مهارات نتيجة تعلمه لموضوع أو مجال دراسي معين. غير أنه يمكن الإفادة من نتائج هذه الاختبارات في انتقاء الطلاب للدراسة في برامج معينة مناسبة لهم، وإخبارهم بإنجازاتهم، وزيادة دافعيتهم، وإثراء تعلمهم. وكذلك في تزويد المعلمين والمسئولين عن بناء المناهج بمعلومات يمكن الاستناد إليها في تطوير هذه المناهج أو تعديلها، وتقييم البرامج التعليمية. كما تُستخدم الاختبارات التحصيلية في التقويم القومي لإجراء مسوح للمعارف والمهارات التي اكتسبها الطلاب من البرامج التعليمية المدرسية في مختلف المجالات الدراسية، وذلك بغرض مراقبة التقدم التربوي على مستوى الوطن. وكذلك تُستخدم هذه الاختبارات لأغراض المساءلة التربوية Accountability للمناطق التعليمية والمعلمين عما تعلمه الطلاب، وفي المنافسات الدولية للتحصيل في الرياضيات والعلوم، وغير ذلك. وسوف أوضح فيما يلي كلاً من هذه الاستخدامات بشيء من التفصيل:

إثراء تعلم الطلاب:

تلعب الاختبارات التحصيلية وبخاصة الاختبارات الصفية التي يكتبها المعلم لطلابه، دوراً رئيساً في عمليتي التعليم والتعلم. فبعض المعلمين يظن أن هذه الاختبارات تطبق عقب تدريس مقرر معين بغرض تقدير درجات كلية للطلاب في هذا المقرر. وعلى الرغم من اهمية بل وضرورة ذلك، إلا أن هذا يعد أحد أدوار الاختبارات التحصيلية؛ فالهدف الرئيس لهذه الاختبارات هو تحسين عملية التعليم، وإثراء عملية التعلم، من خلال التغذية الراجعة التي تقدمها للمعلمين والطلاب فيما يتعلق بفاعلية عملية التعليم، وجودة عملية التعلم. إذ يمكن باستخدام نتائج الاختبار تحديد مدى واقعية وملاءمة الأهداف التعليمية للمواد الدراسية،

ومدى مناسبة طرائق التدريس وتقنياته لمستوى الطلاب. فإذا بينت النتائج أن أداء الطلاب في الاختبار كان ضعيفاً، فإن هذا ربما يدل على أن طريقة التدريس غير مناسبة لهم، أو أن الاختبار لا يعكس الأهداف التعليمية المرجوة، أو لايشتمل على عينة من المفردات ممثلة لهذه الأهداف، أو أن انواع المفردات المستخدمة لا تناسب طبيعة النواتج المحددة.

ولكي تكون التغذية الراجعة فاعلة، ينبغي أن تكون فورية ومستمرة أثناء عمليتي التعليم والتعلم، وأن تفيد الطالب في معرفة اجاباته الصحيحة، وإجاباته الخطأ، لكي يتمكن من تصحيح اخطائه وإثراء تعلمه.

استثارة دافعية التعلم:

فالاختبارات التحصيلية تقيس النواتج التعليمية التي ينبغي أن تكون واضحة للطلاب في مستهل عملية التعليم لكي يبذلوا أقصى جهد لتحقيقها. وعندما تُستخدم هذه الاختبارات في إخبار الطلاب عما حققوه وما لم يحققوه أثناء تعلمهم، فإنه يزيد دافعيتهم للعمل والانجاز. كما أن هذه الاختبارات إذا ما اعتُني بتصميمها وبناؤها بحيث تقيس مهارات معرفية عليا كالتطبيق والتحليل والإبداع، واهتم المعلم بتنمية هذه المهارات المهمة لدى طلابه، فإن الطلاب سوف يبذلون قصارى جهدهم لتعلمها لكي يحصلوا على درجات مرتفعة في الاختبار. ومناقشة نتائج الاختبار مع الطلاب تعد بمثابة تغذية راجعة توضح لهم طبيعة هذه المهارات واهميتها، وكيفية تناول الأسئلة المتعلقة بها. لذلك فإن الاختبارات التحصيلية التي يُعتنى ببنائها يمكن أن تستثير دافعية الطلاب للتحصيل الجيد، وتحقيق الأهداف التعليمية بمختلف مستوياتها. وهذا يساعد في استبقاء المعلومات وتوظيفها وانتقال أثر تعلمها إلى مواقف خارج حدود المنهج المدرسي. كما يساعد في تأكيد أهمية التقويم الذاتي الذي يعد مكونة أساسية من مكونات التعلم المستقل للطالب وزيادة دافعيته، وتقديره لذاته. وبذلك يعتمد الطالب على الضبط والتحكم غير المباشر، والدافعية الداخلية في تحقيق الاهداف التعليمية المرجوة.

انتقاء الطلاب:

تُستخدم الاختبارات التحصيلية في انتقاء الطلاب للدراسة في برامج تعليمية معينة مناسبة، كما أن الطلاب ينتقلون من فرقة دراسية إلى اخرى، ومن مرحلة تعليمية إلى مرحلة

أعلى بناء على نتائجها إضافة إلى نتائج اختبارات استعدادات. لذلك ينبغي أن تتميز هذه الاختبارات بالصدق، وتتميز درجاتها بالثبات.

وكذلك تُستخدم هذه الاختبارات في تسكين الطلاب في الأقسام المختلفة بالكليات، أو في المستويات المختلفة لمقررات معينة. وتُستخدم أيضاً في تحديد ما إذا كان الطالب قد حقق مستوى الأداء المطلوب للقيام بأنشطة معينة.

تقويم المناهج والبرامج التعليمية:

تُستخدم الاختبارات التحصيلية في تقويم مدى ملاءمة المناهج الدراسية، وفاعلية البرامج التعليمية، والمقارنة بين تحصيل الطلاب في المدارس المختلفة. غير أن هذه المقارنة تتطلب الحيطة في تفسير نتائج الاختبارات التحصيلية وبخاصة التي تطبق على نطاق واسع.

المساءلة التربوية:

تُستخدم درجات الطلاب في الاختبارات التحصيلية أيضاً في تقويم أداء المعلم؛ فالمعلم يكون مسؤولاً عن مدى نجاحه في تعليم طلابه. إذ يمكن تطبيق الاختبارات التحصيلية نفسها أو صيغ مكافئة لها قبل قيامه بتدريس مقرر معين، وعقب الانتهاء منه، وذلك لتحديد مكتسبات الطلاب. غير أن هذا الإجراء يشوبه بعض العيوب من الوجهة الإحصائية، حيث إن مقدار ثبات درجات التحسن في أداء الطلاب يكون عادة أقل من ثبات كل من درجات الاختبار القبلي والبعدي.

التقويم القومي:

تُستخدم الاختبارات التحصيلية وغيرها من أدوات تقويم الأداء في إجراء مسوح قومية لمنجزات الطلاب في مختلف المناطق التعليمية. والهدف من ذلك معرفة ما حققه طلاب الوطن من معارف ومهارات واتجاهات في مواد دراسية متعددة مثل: العلوم، والكتابة، والمواطنة، والقراءة، والدراسات الاجتماعية، والرياضيات، والمواد التي تنمي المعارف المهنية، والفنون. ومثال ذلك التقويم القومي للتقدم التربوي في الولايات المتحدة الامريكية الذي يهدف للحصول على معلومات مهمة عن مدى تحقيق النظام التعليمي في مراحل التعليم العام المختلفة للمستويات التربوية Education Standards المتوقعة، وتتبع التغيرات التي تحدث في هذه

المستويات أو النواتج التعليمية عبر الزمن على المستوى القومي، والاستفادة من هذه المعلومات في تحسين صنع القرارات التربوية، وتطوير العملية التعلمية في واقع الصفوف المدرسية.

الاختبارات التحصيلية الصفية التي يكتبها المعلم لطلابه

Teache - Made Classroom Tests:

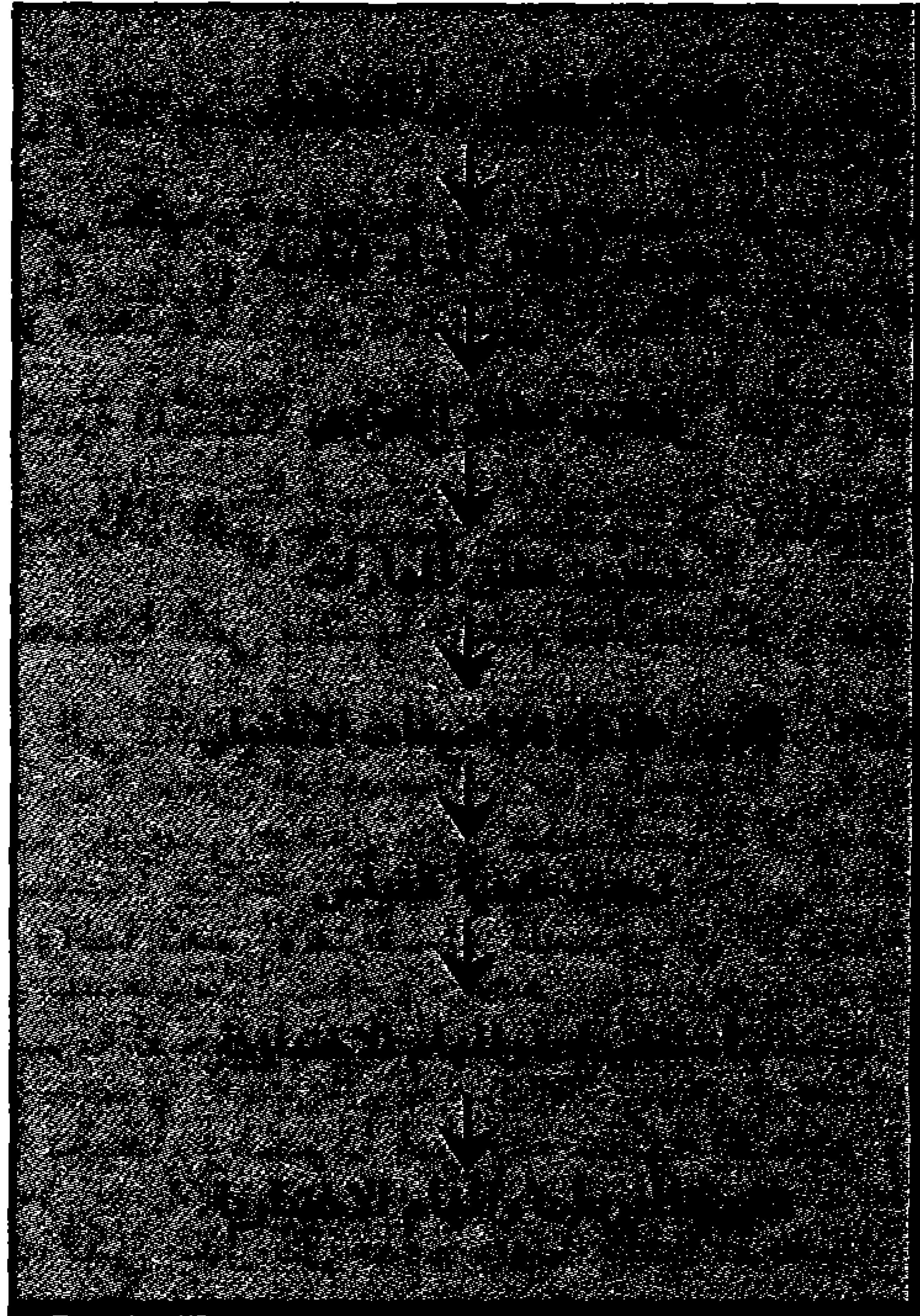
يهتم المعلمون بتقويم طلابهم بصفة دورية متكررة، ويركزون عادة على تقويم أدائهم وتحصيلهم، وذلك لأغراض تعليمية متنوعة كما أوضحنا فيما سبق. ولعل أهم هذه الاغراض من الناحية التربوية هو معرفة مدى تحقيق كل طالب للأهداف أو المستويات المتوقعة Standards للمواد الدراسية المختلفة بغرض تبين مواطن القوة والضعف في تحصيل الطلاب، لكي يتسنى معالجة الصعوبات التي يواجهها معظمهم في تحقيق هذه الأهداف أو المستويات. ولذلك فإن الاختبارات الصفية التي يكتبها المعلم لطلابه تكون عادة مرجعية المحك Criterion -Referenced ، وتستند إلى المنهج الدراسي Curriculum - Based Tests، تناسب حاجات الطلاب الفعلية في صف مدرسي معين. وفي هذه الحالة لا يبغى المعلم مقارنة أداء الطالب بأداء اقرانه في الصف، وإنما تعد الأهداف أو المستويات المتوقعة هي المحك Criterion الذي يتم تقييم الطالب في ضوئه لتحديد درجة تحقيقه هذا المحك. لذلك ينبغي أن يكون المحك واضحاً ومعرفاً تعريفاً دقيقاً، ومحدداً قبل بدء عملية التعليم. وأن يكون الطلاب على دراية تامة بمتطلبات هذا المحك، لكي يبذل كل منهم جهداً كافياً لتحقيقه.

مراحل بناء الاختبارات الصفية:

يتطلب بناء الاختبارات التحصيلية الصفية إعداداً جيداً استناداً إلى خطة تفصيلية مدروسة تراعي الغرض من العملية الاختبارية، وخصائص الاختبارات مرجعية المحك، وطبيعة محتوى المادة الدراسية والمهارات المعرفية المتضمنة في تعلم هذا المحتوى، وأنسب المفردات والمهام الاختبارية التي تقيس هذه المهارات. وهذا يحتاج من المعلم أن يكون متعمقاً في مجاله الدراسي، ومتمكناً من تفاصيله الدقيقة. وأن يكون لديه معرفة كافية وتدرباً جيداً فيما يتعلق بكيفية بناء مختلف أنواع المفردات والمهام التي تقيس الأهداف المعرفية العليا، وصياغة الافكار المتعلقة بالمثيرات التي يمكن أن تشتمل عليها هذه المفردات والمهام صياغة واضحة وموجزة

ودقيقة، أي ترجمة هذه المثيرات إلى مواقف اختبارية مناسبة في ضوء الهدف التعليمي والمحتوى الدراسي. فبناء الاختبارات التحصيلية الصفية يعد علماً وفناً يستطيع معظم المعلمين اتقانه بدرجة معقولة من خلال التدريب المستمر، والخبرة المتراكمة، والممارسة الفعلية.

ويوضح الشكل التخطيطي (5 - 2) التالية المراحل الرئيسة في اعداد هذه الاختبارات:



شكل (5 - 2) يوضح خطوات بناء الاختبارات التحصيلية الصفية

وسوف أوضح كل من هذه الخطوات بشيء من التفصيل فيما يلي:

(أولاً) تحديد الغرض من الاختبار:

قبل البدء في تصميم وبناء الاختبار التحصيلي الصفي ينبغي على المعلم أن يحدد الغرض الرئيس الذي من أجله يعد الاختبار، وذلك لأن هذا الغرض هو الذي يوجه جميع الخطوات التالية. فقد سبق أن اوضحنا أن الاختبارات التحصيلية الصفية تتعدد أغراضها وتتنوع استخداماتها بحسب اهداف عملية التقويم، أي ما إذا كان التقويم قبلياً تمهيدياً لمعرفة

المستوى المدخلي للطلاب، أو تشخيصياً لتعرف جوانب القوة والضعف في أداء الطلاب أثناء عمليتي التعليم والتعلم، أو ختامياً لوضع تقديرات للطلاب عقب الانتهاء من دراسة وحدة تعليمية أو مقرر بأكمله. لذلك ينبغي على المعلم أن يحدد في ضوء ذلك ما يلي:

(أ) الغرض من اختبار طلابه، ومتى سيختبرهم.

(ب) فيم تُستخدم نتائج الاختبار، ونوع القرارات التي ستتخذ في ضوء هذه النتائج.

(جـ) المرحلة التعليمية، والصف الدراسي، والمستوى العمري للطلاب.

(د) المادة أو الوحدة الدراسية، أو الموضوعات التي يود قياس التحصيل فيها.

(هـ) الزمن المتاح لتطبيق الاختبار، وما تتطلبه الإجابة من أدوات أو مصادر.

فمما لا شك فيه أن تحديد هذه الأمور تحديداً واضحاً يُمكن المعلم من تصميم اختباره تصميماً مناسباً يحقق الغرض الرئيس منه، ويفيد من نتائجه الفائدة المرجوة.

(ثانياً) تحديد المجال أو النطاق المراد قياسه:

إن تحديد المجال أو النطاق المراد قياسه يعد خطوة اساسية في بناء الاختبار التحصيلي الصفي مهما اختلف الغرض منه. فبعد أن يحدد المعلم الغرض الرئيس من الاختبار، أي لماذا يختبر طلابه، وفيم تُستخدم نتائج الاختبار، ينبغي معرفة ماذا يقيس الاختبار.

فقياس التحصيل الدراسي شأنه شأن قياس الظواهر المختلفة يتطلب تحديد المتغير المراد قياسه، نظراً لأن الاختبار باعتباره أداة قياس ما هو إلا ترجمة إجرائية لهذا المتغير. وكلما كان التحديد دقيقاً لا يحتمل الجدل أو التأويل كان محتوى الاختبار أكثر مواءمة ووضوحاً. وتعريف هذا المتغير يكون في ضوء نطاق سلوكي Behavioral Domain معرفاً تعريفاً إجرائياً. ويختلف خبراء القياس والتقويم في تعريفهم لمفهوم النطاق Domain؛ فمنهم من يرى أنه يحتوي الأهداف الضمنية أو الصريحة أو الموضوعات المراد قياسها، أو أنه مخطط للمعارف والقدرات التي يشتمل عليها الاختبار، وعدد المفردات التي تقيس أنماط المعارف المختلفة، وكيفية انتقاء النوع المناسب من المفردات والمهام الاختبارية، على أن يتضمن المخطط ما ينبغي إجراؤه على المعارف، مثل: فهم المصطلحات والحقائق والتعميمات، التنبؤ بالنواتج،

تقديم تفسيرات، حل مشكلات، وهكذا. ومنهم من يرى البدء بجدول مواصفات للاختبار كوسيلة تنظيمية لتصميم خطة بناء المفردات والمهام بحيث تضمن توازن المحتوى، ومستويات السلوك أو العمليات المعرفية التي تتطلبها الإجابة على المفردات والمهام الاختبارية. كما يرى بعض آخر أن تحديد النطاق السلوكي يتطلب إعداد جدول المواصفات، إضافة إلى توصيف المفردات والمهام.

ويمكن القول بعامة أن النطاق Domain هو مصطلح مستمد من العلوم الفيزيائية، ويقصد به في هذه العلوم مجموع المعلومات التي يتوقع أن تفسرها نظرية معينة. وقد استخدم خبراء القياس هذا المصطلح للإشارة إلى النطاق الشامل للمحتوى Content Universe الذي يفترض أن الاختبار يشتمل على مفردات أو مهام تمثل عينة مستمدة من هذا النطاق الشامل. وتتطلب معاينة النطاق Domain Sampling تحديد مكوناته وعناصره، وهذا ما يجعل تحديد النطاق خطوة أساسية في تصميم خطة الاختبار.

وعادة يحدد النطاق السلوكي للاختبارات التحصيلية الصفية بأهداف سلوكية أو نواتج متوقعة للمادة الدراسية. وهذا يتطلب معرفة ما يجب تعلمه في مادة دراسية معينة، وكيف يتم تعليمها وتعلمها قبل البدء في بناء الاختبار. لذلك لا يجب أن تكون الأهداف التعليمية مجرد قوائم لمهارات غير منظمة، ومهام بسيطة محدودة الفائدة. والطريقة الاكثر فاعلية في هذا الشأن هي البدء بالمفاهيم أو المهام الرئيسة للطلاب، ومن ثم تحديد المحتوى والمهارات المتعلقة بهذه المفاهيم أو المهام (MSEB, 1991).

(ثالثاً) تحديد نطاق المحتوى:

عند تحليل نطاق المحتوى Content Domain يفضل الاسترشاد بأحد الأطر المرجعية التي حاولت تصنيف محتويات المواد الدراسية ونواتجها التعليمية، مثل: تصنيف بلوم المعدل Bloom's Taxonomy، أو تصنيف موراي، ووارد Murray & Ward Taxonomy، أو غيرهما من التصنيفات. ويمكن تلخيص ذلك في جدول (5 - 1) التالي (Ward & Ward, 1999) :

تصنيف بلوم	تصنيف موراي. induced
	العناصر
(1) النوعيات	(أ) المصطلحات
المصطلحات	(ب) الحقائق
الحقائق	(جـ) الاتفاقات
	(د) الرموز
	(هـ) المكونات
(2) الطرق والأساليب	الطرق والأساليب (التفاعلات)
الاتفاقات	(و) النزعات
التصنيفات والأقسام	(ز) نتائج البحوث
النزعات والتتابعات	(ح) القواعد والقوانين
المنهجية	(ط) الأساليب
المحكات	(ك) المحكات
	العموميات
(3) العموميات والتجريدات	(ل) المبادئ
المبادئ والتعميمات	(م) النظريات
النظريات والبنى	(ن) المعايير والأخلاقيات

جدول (2 - 1) يوضح تصنيف المحتوى عند كل من بلوم، وموراي ووارد.

ولتحديد نطاق المحتوى ينبغي تصميم خطة المحتوى Content Outline التي تنظم المحتوى المراد قياسه في وحدات منطقية، وتحديد أوزان أو نسب مئوية لكل منها. ويمكن اتباع الخطوتين الرئيستين التاليتين في هذا الشأن:

(أ) تحديد المتغير المراد قياسه: ونقصد بالمتغير المفهوم، أو المبدأ، أو الصيغة، أو الخاصة المراد قياسها. ويمكن أن يكون المتغير متسعاً، مثل: "المبادئ والقوانين العلمية"، أو محدوداً مثل: "الأعداد المحصورة بين صفر، 10". وتعريف المتغير يعد بمثابة الإطار المرجعي الذي يُستند اليه في تحديد نطاق المحتوى الذي سيتناوله الاختبار. وهذا يعتمد على الهدف من

استخدام نتائج الاختبار؛ فإذا كان الهدف توجيه مسار عملية التعليم، أي التقويم البنائي، فإنه يفضل أن يكون المتغير محدوداً، اما إذا كان الهدف هو التقويم الختامي، فإن هذا ربما يتطلب أن يكون المتغير متسعاً. ويمكن الاعتماد في تحديد المتغير المراد قياسه على مفردات المناهج، أو محتويات الكتب المدرسية، أو أي مصادر معلومات اخرى مناسبة.

(ب) تحديد المحتوى وتنظيمه: والهدف من ذلك التحقق من عدم وجود تداخل في المحتوى، حيث يمكن أن يتضمن موضوعاً معيناً جميع الموضوعات الأخرى. لذلك يفضل تحديد الموضوعات الرئيسة وما يندرج تحتها من موضوعات فرعية مثل:

جغرافية أقطار الوطن العربي في إفريقيا.

- جغرافية ليبيا والسودان.
- جغرافية أقطار المغرب العربي.
- جغرافية جمهوريتي موريتانيا الإسلامية والصومال.

كما يمكن الاسترشاد بالتصنيف الموضح بجدول (5 - 1) السابق في تحديد مكونات محتوى وحدة دراسية معينة، مثل: العناصر وفرعياتها، الطرق والأساليب وفرعياتها، العموميات وفرعياتها، وذلك في ضوء الموضوعات التي تشتمل عليها هذه الوحدة. ويمكن تنظيم المحتوى ومكوناته في جدول مقسم بحسب عدد الموضوعات المعينة.

(جـ) تحديد أوزان المحتوى: بعد تحديد الموضوعات وتنظيمها، ينبغي تقدير أوزان لكل من الموضوعات الرئيسة والفرعية، من حيث أهميتها النسبية في الوحدة أو المادة الدراسية. وهذا يساعد المعلم في إمعان التفكير في مكونات مادته الدراسية وعناصرها والمطلوب تعلمه، والزمن الذي يستغرقه تدريس الموضوعات المختلفة، والجهد اللازم لذلك. ويمكن التعبير عن هذه الأوزان بنسب مئوية لكل من الموضوعات الرئيسة والفرعية.

(رابعاً) تحديد نطاق المهارات:

بعد تحديد نطاق المحتوى تكون الخطوة التالية تحديد نطاق المهارات Skills Domain التي سوف يقيسها الاختبار في ضوء هذا المحتوى، وكذلك أوزانها النسبية.

وتحديد نطاق المهارات يتطلب معرفة العمليات العقلية التي يمكن أن يستخدمها الطالب في

إجابة مفردة أو مهمة اختبارية معينة تتعلق بأحد المحتويات. وتتباين الموضوعات في عدد المهارات التي تتطلبها ومستوياتها المعرفية، فبعضها ربما يتطلب مهارة واحدة من مستوى معرفي مرتفع، والبعض الآخر ربما يتطلب عدة مهارات تختلف في مستواها المعرفي. وفي أي الحالات ينبغي صياغة المهارة صياغة واضحة ومحددة، بحيث يمكن ملاحظة السلوك المتعلق بها، وقياسه بمفردات أو مهام اختبارية.

ويمكن الاسترشاد أيضاً في تحديد نطاق المهارات بالتصنيفات التي أشرنا إليها فيما سبق. ويوضح جدول (2 - 2) التالي تصنيفات المهارات عند كل من بلومBloom، وموراي ووارد Ward & Ward 1999)

تصنيف بلوم	تصنيف موري. ووارد
(1) المعرفة (التذكر) (2) الفهم - الترجمة - التفسير - الاستكمال (3) التطبيق (4) التحليل (5) التركيب (6) التقويم	(1) التعرف (2) الاستدلال - الترجمة (التحويل، التصنيف) - التفسير (التلخيص، التعميم) - الاستكمال (العلية، التنبؤ، التضمينات) (3) التحليل (التشابهات، المصفوفات، التتابع، المخططات، الحقائق، الافتراضات، الفروض، الآراء، الاتساق الداخلي، وجهات النظر) (4) التقويم - المقارنة بمعايير محددة - المقارنة بمعايير غير محددة - الإسهام في غرض أو تأثير معين - الموازنة / المغايرة (5) حل المشكلات - تحديد المشكلات - اختيار المبدأ المناسب - استخدام المبدأ الصحيح الى [illegible]

جدول (5 - 2) يوضح تصنيف المهارات عند بلوم، وموراي ووارد

ويعد تحديد نطاق المهارات امراً مهماً لأنه يُستند إليه في تحديد المستوى المعرفي أو المهاري للمفردات أو المهام الاختبارية. لذلك ينبغي أن يكون السلوك المراد ملاحظته وقياسه لدى الطالب واضحاً لكي يتمكن المعلم من التحقق من أداء الطالب للمهارة المرجوة. وهذا يتطلب تعريف المتغيرات أو المهارات في ضوء الإجراءات التي سوف تُستخدم في قياسها. ولعل أفضل الطرق لتعريف المهارات تعريفاً اجرائياً هو إعداد مواصفات للمفردات والمهام التي سوف يستند اليها في قياس هذه المهارات في ضوء نطاق المحتوى الذي تم تحديده. وهذا يعني تحديد المثيرات التي سوف تشتمل عليها المفردات أو المهام الاختبارية، وخصائص الإجابة على كل منها.

ويمكن اتباع الخطوتين التاليتين لتحديد نطاق المهارات:

(أ) تحديد المهارات المتوقع ارتباطها بكل موضوع من موضوعات المحتوى: وتتطلب هذه الخطوة وضوح نطاق كل من المحتوى والأهداف، والاسترشاد في ذلك بجدولي (5 - 1)، (5 - 2) السابقين.وعندئذ يمكن تحديد أنواع المهارات ومستوياتها المناسبة لكل موضوع فرعي، حيث إن الموضوعات المختلفة ربما تتطلب مستويات مهارية متباينة. فمهارة التحليل تتطلب مستويات أخرى من المهارات يمكن قياسها؛ فتحليل قطعة مقروءة ربما يتطلب معرفة الكلمات، وفهم معناها، والاستدلال، والربط، وإعادة الصياغة، وغير ذلك.

(ب) اختيار المهارات المناسبة: يمكن قياس مهارة واحدة في موضوع معين، كما يمكن قياس عدة مهارات في الموضوع نفسه. وهذا يعتمد على الهدف من الاختبار، وتوقعات المعلم عن مستوى المهارات التي يرى ضرورة اكتسابها، وعدد مفردات الاختبار، والزمن المتاح للإجابة عليه، وأنواع المفردات والمهام المناسبة للمحتوى والمهارات المراد قياسها.

(خامساً) تكوين جدول مواصفات الاختبار:

بعد تحديد نطاق كل من المحتوى، والمهارات والأوزان التي تدل على الأهمية النسبية لكل من الموضوعات الرئيسة والفرعية، والمهارات ذات المستويات المختلفة، فإن المعلم يحتاج إلى وسيلة تنظيمية تساعده في ربط المهارات بالمحتوى بحيث يمكن تقدير أوزان تتناسب مع أهمية

المهارة والمحتوى في آن واحد لكي يضمن توازن جميع مستويات المهارات في اطار محتوى المادة الدراسية، وتمثيلها في الاختبار. وعلى أساس هذه الأوزان يمكنه تحديد عدد المفردات والمهام الاختبارية المطلوب اعدادها. وهذه الوسيلة التنظيمية تسمى «جدول مواصفات الاختبار Test Specification Table».

وهذا الجدول ثنائي البعد، حيث يمثل بُعده الأفقي مستويات المهارات ويمثل بُعده الرأسي اقسام المحتوى (الموضوعات الرئيسة والفرعية)، ومن ثم تحدد في خلايا هذا الجدول الأوزان المتعلقة بالأهمية النسبية لكل من المهارة والمحتوى باستخدام أوزان المحتوى، وأوزان المهارة التي سبق تحديدها في الخطوتين الثالثة والرابعة. وباستخدام هذه الأوزان يمكن تحديد عدد المفردات والمهام التي يجب إعدادها لكل مهارة في إطار كل موضوع من الموضوعات المحددة. ويمكن إجراء ذلك بأن نضرب الوزن النسبي (النسبة المئوية) للعمود الذي تقع فيه خانة معينة في الجدول × الوزن النسبي للصف الذي تقع فيه الخانة نفسها، ثم نضرب الناتج × عدد المفردات والمهام الكلية التي سوف يشتمل عليها الاختبار.

(سادساً) تحديد كيفية القياس:

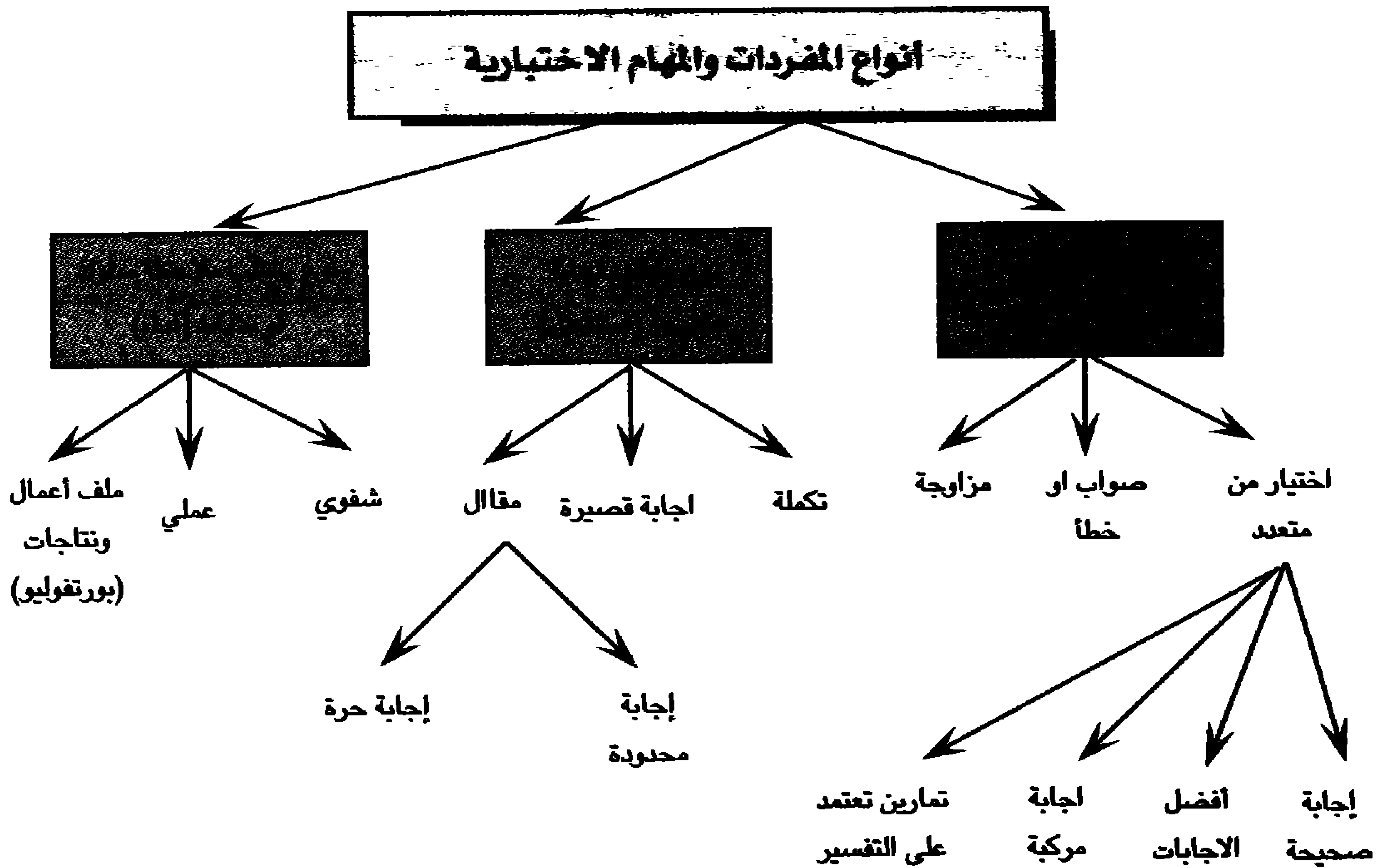
يتطلب قياس المهارات التي تم تحديدها في ضوء نطاق المحتوى أنواعاً مختلفة من المفردات والمهام الاختبارية. وكثير من المهارات تتطلب مستويات معرفية عليا، وأداءات، وعمليات، ونتاجات. لذلك فإنه ينبغي أن يكون المعلم على دراية بفنيات كتابة مختلف أنواع المفردات الاختبارية التي تتطلب الورقة والقلم، وكذلك المهام الاختبارية الأخرى التي تتعلق بأساليب ملاحظة السلوك، وموازين التقدير، وملفات الأعمال (البورتفوليو)، وتقييم النتاجات، وغير ذلك. وقد سبق أن أوضحنا في مستهل هذا الفصل تصنيف الاختبارات التحصيلية وفقاً لأنواع مفرداتها ومتطلبات إجابتها.

ويمكن القاء الضوء على بعض من هذه المفردات والمهام، حيث إن كثيراً من المهارات العليا المهمة يصعب قياسها بمفردات اختيار من متعدد، أو صواب أو خطأ، أو مزاوجة، أو إكمال. فمثلاً مهارة كتابة تقرير علمي استناداً إلى بيانات احصائية يحتاج قياسها إلى سؤال مقال مفتوح وفق محكات معينة، ومهارة الإلقاء أو التعبير الشفوي تتطلب اختباراً شفوياً، ومهارة إجراء تجربة مختبرية تتطلب أداء عملياً، ومهارة تصميم وتنفيذ مشروع فردي أو جماعي معين

تعد من المهارات الموسعة التي يمكن تقييمها استناداً إلى ملف أعمال Portfolio، وهكذا.
ويعتمد انتقاء نوع المفردات أو المهام الاختبارية على عوامل متعددة مثل: الغرض من تقييم الطلاب، والموارد المتاحة، وكيفية قياس المهارات الرئيسة بمفردات ومهام مناسبة بدرجة أفضل، وعدد الطلاب المختبرين، وتدرب المعلم على كتابة مفردات تقيس مهارات معرفية عليا، ومهارات ادائية وعملية وشفوية، وطريقة التصحيح وتأثرها بالجوانب الذاتية، وغير ذلك.

(سابعاً) بناء المفردات والمهام الاختبارية:

نظراً لتعدد وتنوع المفردات والمهام التي يمكن أن يشتمل عليها الاختبار التحصيلي الصفي، فإنه يمكن تصنيفها في الشكل التخطيطي (5-3) التالي:



شكل (5 - 3) يوضح الأنواع المختلفة للمفردات والمهام الاختبارية في المجال المعرفي

ويتضح من شكل (5 - 3) أن المفردات والمهام الاختبارية يمكن أن تنقسم إلى ثلاثة أنواع رئيسة كالتالي:

(1) نوع يتطلب إجابة مقيدة:

وفيه يختار الطالب إجابته من بين بدائل إجابات معطاة، حيث أن الإجابة تعتمد على عملية التعرف Recognition. ومن هذا النوع ما يلي:

(أ) مفردات الاختيار من متعدد بصورها المختلفة: وتتطلب إجابة صحيحة، أو افضل الإجابات، أو إجابة مركبة، أو تفسيرات لبيانات ومعلومات معطاة، وتعد مفردات الاختيار من متعدد اكثر مفردات هذا النوع استخداماً وملاءمة لقياس معارف ومهارات تعليمية متنوعة من مستويات معرفية مختلفة، كما تناسب طلاب جميع المراحل التعليمية.

(ب) مفردات الصواب أو الخطأ: وتتطلب تحديد صحة أو عدم صحة عبارة معينة، أو الحكم على صدق أو عدم صدق قضية محددة، أو التمييز بين الحقائق والآراء، وغير ذلك من الإجابات التي تتطلب الاختيار بين بديلين.

(جـ) مفردات المُزاوجة: وتتطلب الربط بين الاحداث، والتعرف على العلاقات البسيطة بين الأشياء أو الظواهر، أو تصنيفها في اقسام معينة.

وتعد جميع هذه الاشكال من المفردات التي تتطلب إجابة مقيدة مفردات موضوعية، من حيث أن تصحيحها لا تتدخل فيه العوامل الذاتية للمعلم أو القائم بالتصحيح، ودرجاتها ثنائية، اما واحد صحيح أو صفر.

(2) نوع يتطلب إجابة مفتوحة:

وفيه يجيب الطالب كتابة بأسلوبه الخاص ما يراه مناسباً للسؤال، حيث أن الإجابة تعتمد على عملية الاستدعاء Recall وليس التعرف Recognition. ويتفاوت طول الإجابة ومتطلباتها بحسب نوع المفردة أو السؤال. ومن هذا النوع:

(أ) فقرات الإكمال والإجابة القصيرة: وتتطلب اكمال عبارة بكلمة أو كلمتين، أو يمكن في مفردات الإجابة القصيرة أن يجيب الطالب إجابة محدودة بجملة أو جملتين. وهذه المفردات الاختبارية تناسب قياس تذكر المصطلحات والحقائق، وتحديد العلاقات والأسباب، وصياغة الفروض، وغير ذلك من المهارات البسيطة.

(ب) أسئلة المقال: ويتفاوت طول الإجابة في هذه الأسئلة بحسب طبيعة السؤال، إذ يمكن أن يتطلب سؤال المقال Essay Question إجابة محدودة، ولكنها تكون عادة اطول مما تتطلبه مفردات الإجابة القصيرة. أو ربما يتطلب إجابة مستفيضة مثل: كتابة مقال، أو تعليق حول أحداث معينة، أو نقد قضية، أو تبرير موقف معين، أو كتابة خطة بحث، وهكذا. وتعد

اسئلة المقال أكثر أنواع المفردات الاختبارية استخداماً في قياس المهارات المعرفية العليا كالتطبيق، والتحليل، والتركيب، والتقويم، وبخاصة تلك التي تتطلب التعبير الحر، وتنظيم المعلومات، ونقد الأفكار، واقتراح البدائل، وحل المشكلات، والمسائل الرياضية التي تحتاج خطوات متعددة، وغير ذلك من المهارات العقلية المهمة.

(3) نوع يتطلب ملاحظة سلوك أو نشاط: يتطلب كثير من المواد الدراسية قيام الطالب بأنشطة وأداءات متنوعة بعضها يتطلب عمليات معينة وتؤدي إلى نتاجات محددة أو اصيلة. ولعل اهتمام خبراء القياس والتقويم بهذه الأداءات والنتاجات قد تزايد بدرجة ملحوظة في الآونة الأخيرة، بعد ما تبين أن مفردات النوع الأول التي تتطلب إجابة مقيدة تكون اصطناعية، وتقيس السلوك بطريقة غير مباشرة، وتحد من تفكير الطالب وإبداعه. لذلك اتجه هؤلاء الخبراء إلى ما اصبح يسمى التقويم الواقعي أو البديل Authentic Assessment الذي يؤكد تقييم الأداءات والعمليات والنتاجات الفعلية التي تتميز بالواقعية، والاصالة، والقياس المباشر، وهو ما تؤكده المناهج المدرسية. وتقييم الأداءات والنتاجات يكون اكثر فاعلية، ويقدم اكبر قدر من المعلومات عن التحصيل الحقيقي للطالب إذا اعتني بتصميم مهام الأداء Performance Tasks. وتقدير الدرجات باستخدام موازين تقدير وصفية Rubrics، بحيث تكون نتائجها موضوعية وموثوق بها. ولا يختلف إعداد مهام الأداء Performance Tasks عن مهام النتاجات Production Tasks إلا في موعد إجراء التصحيح. فتصحيح مهام الأداء التي تتطلب القيام بعمل معين، مثل: توصيل دائرة كهربائية، أو تصميم نموذج هندسي، أو رسم خريطة جغرافية، يحتاج إلى مراجعة ابعاد ومكونات العمل ومطابقته للمواصفات المطلوبة. ويمكن تقييم الأداء عن طريق تحليل العمل أو المهمة المعينة إلى خطوات أو عمليات ينبغي أن يقوم بها الطالب اثناء الأداء أو تنفيذ العمل. ثم تدون نتائج هذا التحليل على بطاقة ملاحظة يسجل عليها المعلم تقديراته لأداء الطالب في كل خطوة أو عملية اثناء قيامه بتنفيذ العمل. لذلك ينبغي العناية باعداد بطاقة الملاحظة أو ميزان التقدير بحيث يكون تحليلياً موضحاً العمليات أو العناصر التي يشتمل عليها العمل أو المهمة الادائية، ويفضل عرضها على الطالب قبل قيامه بالمهمة لكي يتفهم هذه العناصر، ويقوم بالمراقبة الذاتية لأدائه اثناء تنفيذ المهمة، مما يزيد من جودة هذا الأداء. وينطبق ذلك أيضاً على المهام الشفوية والعملية.

أما مهارات النتاجات فإنه يتم تصحيحها وتقدير درجاتها عقب انتهاء الطالب من العمل وتوصله للنتاج المطلوب. والمحك الذي يستند إليه تقدير الدرجات في هذه الحالة هو مدى جودة هذا النتاج، أو مدى صحة النتيجة التي وصل إليها الطالب.

ويفيد الأسلوب الأول في تشخيص جوانب الضعف في أداء الطالب أثناء قيامه بعمل أو مهمة معينة، مما يساعد في توجيهه ومعالجة أخطائه. بينما يفيد الأسلوب الثاني في تقدير المهارة أو النتاج النهائي في المواقف الجماعية في المختبر، أو حجرة الرسم، أو الملعب في وقت قصير نسبياً.

وسوف أتناول فيما يلي مزايا وعيوب كل من هذ الأنواع، وكيفية بنائها. وقد سبق للمؤلف أن تناول ذلك بمزيد من التفصيل في مختلف المجالات الدراسية في عدة مراجع سابقة يمكن للقارئ المهتم الرجوع اليها والاسترشاد بها في هذا الشأن، ولعل أهمها: دليل المعلم في بناء وتصميم الاختبارات التحصيلية الصفية المعاصرة، ودليل المعلم في تقويم الطلبة في الدراسات الاجتماعية، والاختبارات التشخيصية مرجعية المحك، والتقويم التربوي البديل.

ويمكن تلخيص مزايا وعيوب النوعين الرئيسين للمفردات الاختبارية، وهما النوع الذي يتطلب إجابة مقيدة، والنوع الذي يتطلب إجابة مفتوحة في الشكل (2 - 4) التالي:

مقارنة بين النوعين الرئيسين للمفردات الاختبارية

عنصر المقارنة	مفردات تتطلب إجابة مقيدة (مفردات موضوعية)	مفردات تتطلب إجابة مفتوحة (أسئلة المقال)
(1) الأهداف التعليمية المناسبة	تتطلب أن يختار الطالب إجابة صحيحة من بين مجموعة من البدائل، أو يقدم إجابة محدودة بكلمة أو كلمتين أو جملة قصيرة، لذلك فهي تقيس بدرجة أفضل تذكر وفهم الحقائق والمعارف والمبادئ، والتعميمات، وكذلك يمكن أن تقيس مستويات التطبيق والتحليل والتقويم.	تتطلب أن يعبر الطالب بنفسه مستخدماً ما في حوزته من معلومات وأفكار، لذلك فهي تناسب قياس قدرة الطالب على تنظيم معلوماته وأفكاره، وتركيبها في نسق متكامل وكذلك تناسب قياس القدرة على الاستنتاج السليم، وعمليات التفكير ونواتجه، ولا تقيس تذكر المعلومات بفاعلية.
(2) شمول المحتوى	تتناول نطاقاً واسعاً من المعارف في اختبار ما نظراً لإمكانية الإجابة عن عدد كبير من هذه الفقرات في وقت قصير نسبياً.	تتناول نطاقاً محدوداً من المعارف في اختبار المقال، نظراً لما يتطلبه السؤال من إجابة تفصيلية تستغرق وقتاً طويلاً نسبياً.

(3) مدى سهولة الإعداد	تستغرق وقتاً طويلاً في إعدادها وبخاصة مفردات الاختيار من متعدد، وتتطلب فكراً مبدعاً من جانب المعلم وجهداً ملحوظاً.	تستغرق وقتاً أقل في إعدادها. غير أنها تتطلب جهداً في تحديد الإطار المرجعي للسؤال، وإعداد نموذج مناسب لتصحيحه.
(4) تصحيح الإجابات	يمكن تصحيحها بيسر وسهولة بواسطة مفتاح تصحيح او بطريقة آلية، حيث أن الاجابة اما أن تكون صحيحة أو خطأ، والتصحيح يكون موضوعياً، اي لا تتدخل فيه عوامل ذاتية.	تستغرق وقتاً وجهداً في تصحيحها وبخاصة اذا كان عدد الطلاب كبيراً، كما أن عملية التصحيح تتأثر بعوامل ذاتية متعددة إذا لم يكن هناك نموذج جيد يسترشد به في التصحيح.
(5) الثبات والصدق	نظراً لاحتواء الاختبار عادة على عدد كبير من هذه الفقرات، فإن الاختبار يكون اكثر صدقاً. ونظراً لموضوعية التصحيح فإن الدرجات تكون اكثر اتساقاً او ثباتاً.	نظراً لاحتواء الاختبار عادة على عدد قليل نسبياً من الأسئلة، فإن محتوى الاختبار يكون اقل صدقاً، ونظراً لتدخل عوامل ذاتية في عملية التصحيح، فإن ثبات درجاته يكون منخفضاً في كثير من الأحوال.
(6) إثارة دافعية الطلبة	تشجع الطلبة على تأسيس قاعدة عريضة من المعارف والمهارات.	تشجع الطلبة على تنظيم الأفكار والتعبير عنها بفاعلية.
(7) القيمة التشخيصية	تفيد فائدة كبيرة في تشخيص صعوبات التعلم، وجوانب الضعف وفجوات التحصيل.	قليلة الفائدة في التشخيص، إلا إذا اهتم المعلم بالعمليات التي استخدمها الطالب في التوصل إلى النواتج.
(8) اثر التخمين في الإجابات	تتأثر إجاباتها بعامل التخمين، وبخاصة فقرات الصواب او الخطأ التي تتطلب الاختيار بين بديلين، حيث تصل قيمة احتمال التخمين إلى 0.50، غير انه يمكن بطرق معينة التغلب على ذلك إلى	لا تتأثر إجاباتها بعامل التخمين لأنه لا توجد بدائل يختار الطالب من بينها، وإنما يستدعي معلوماته من ذاكرته بنفسه ويوظفها بأسلوبه الخاص في إطار المطلوب من السؤال.

شكل (5 - 4) يوضح مزايا وعيوب الأنواع الرئيسة للمفردات الاختبارية

كيفية بناء مفردات الاختيار من متعدد:

تعد مفردات الاختيار من متعدد أكثر أنواع المفردات التي تتطلب إجابة مقيدة استخداماً، وبخاصة في الاختبار التحصيلية المقننة.

وهناك العديد من الدراسات التي اهتمت بتعرف مدى فاعلية هذا النوع من المفردات بالمقارنة بأسئلة المقال التي تتطلب إجابة مفتوحة، وكذلك أفضل الطرق لكتابة هذه المفردات.

ويتميز هذا النوع من المفردات بالمرونة في طريقة استخدامها في قياس المستويات المعرفية المختلفة عن غيرها من أنواع المفردات الاخرى، كما يمكن استخدامها في مختلف المجالات الدراسية، وفي جميع مراحل التعليم.

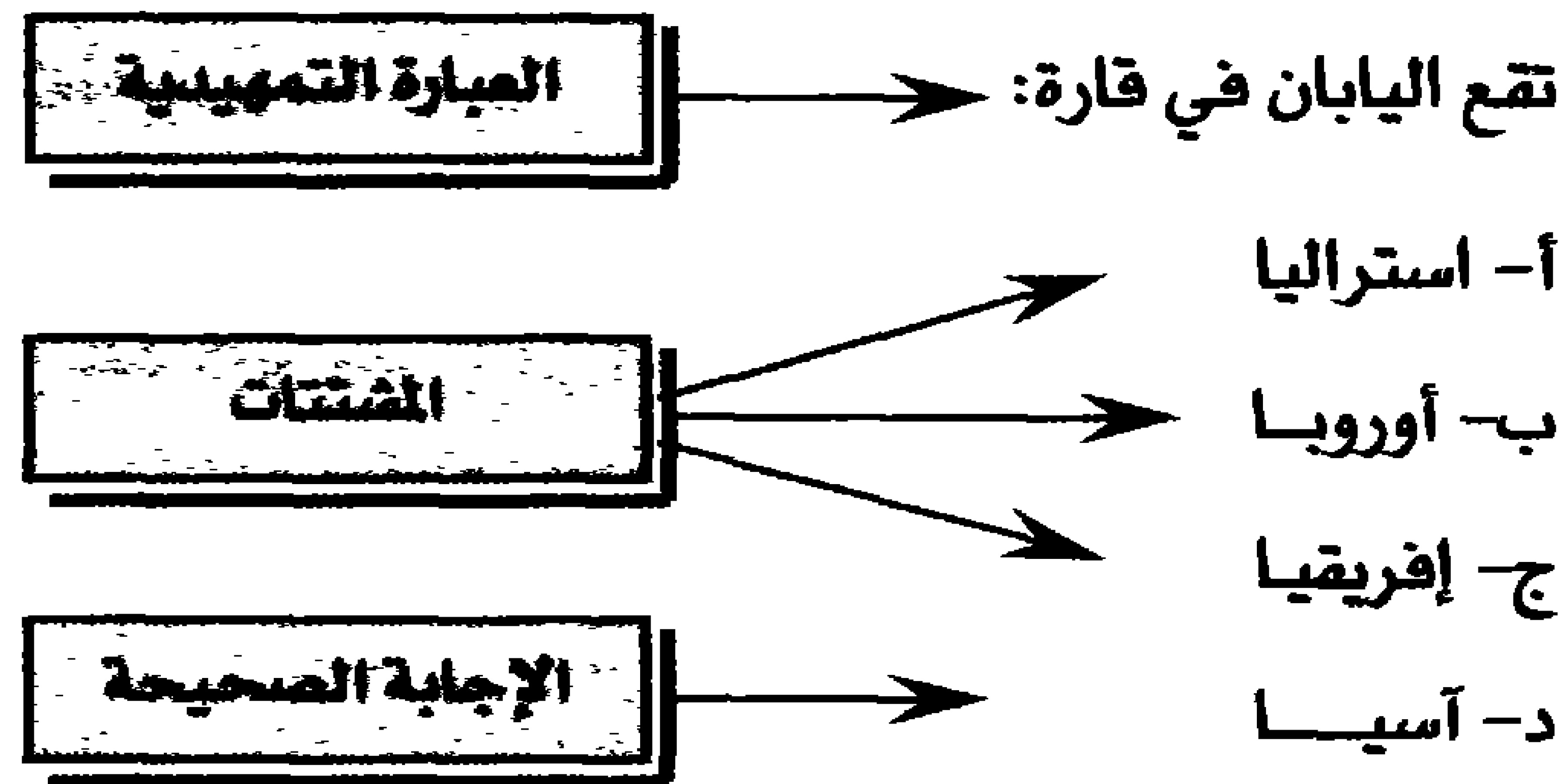
ونظراً لانه يمكن إجابة عدد كبير من هذه المفردات في مدة قصيرة، فإن اختبار الإختيار من متعدد يمكن أن يشتمل على عدد كبير نسبيا من المفردات التي تقيس مختلف الأهداف أو المستويات التعليمية في كثير من الموضوعات، وبذلك تكون عينة مفردات الاختبار ممثلة للأهداف أو المستويات في إطار محتوى المجال الدراسي المعين، كما أنه يمكم تصحيحها بسرعة وموضوعية.

غير أن بناء مفردات الاختيار من متعدد وبخاصة التي تقيس المستويات المعرفية العليا تتطلب تدرباً كبيراً ومهارة فائقة ودراية عميقة بالمجال الدراسي الذي تقيسه هذه المفردات، وكذلك مهارة في استخدام اللغة، كما أنها تستغرق وقتاً طويلا في كتابتها. ولكن إذا وضعنا في اعتبارنا أنها تستغرق وقتا اقل في تصحيحها وبدرجة أكثر موضوعية من الأسئلة التي تتطلب إجابة حرة، وامكانية استخدام هذه المفردات مرات متتالية مع مجموعات أخرى من الطلاب لوجدنا انها أكثر فعالية من أسئلة المقال مثلاً في قياس الأهداف أو المستويات المناسبة لها.

وتتكون مفردة الاختيار من متعدد من عبارة تمهيدية، أو مشكلة ما، أو سؤال stem، وقائمة مقترحة من الحلول لهذه المشكلة، أو مجموعة من البدائل يكون عددها عادة أربعة أو خمسة أحدها يمثل الإجابة الصحيحة أو أفضل الإجابات، وتسمى هذه البدائل المشتتات distractors، وتُكتب البدائل غير الصحيحة بحيث تبدو في ظاهرها للطالب غير المستذكر أنها صحيحة بينما هي في الحقيقة إجابة خطأ. ووظيفة هذه المشتتات التمييز بين الطلاب الذين لا يعرفون الإجابة الصحيحة، أو لديهم معلومات جزئية عن حل المشكلة الموضحة في العبارة التمهيدية، لذلك يختار الطالب غير المستذكر أو الذي لم يتقن المادة المطلوبة أحد هذه المشتتات دون الإجابة الصحيحة.

ويعتمد استخدام عبارة أو جملة ناقصة أو سؤال في صياغة المشكلة على عدة عوامل، منها: أن السؤال المباشر يعد أسهل في كتابته، ويناسب تلاميذ المرحلة الابتدائية، ويمكن عن طريقه عرض المشكلة بصورة واضحة. أما الجملة الناقصة إذا أُحسن صياغتها، فإنها يمكن أن توضح المشكلة بصورة أكثر دقة. ويُنصح الشخص المبتدأ بكتابة المشكلة في صورة سؤال مباشر ثم يُحوله إلى جملة ناقصة إذا وجد أن المشكلة ستكون أكثر وضوحاً ودقة بهذا الشكل الجديد.

ويمكن توضيح مكونات مفردة الاختيار من متعدد على المفردة البسيطة التالية:



تقع اليابان في قارة:

أ- استراليا

ب- أوروبا

ج- إفريقيا

د- آسيا

ومن الأفضل عند اختيار هذه المشتتات استخدام الأخطاء الشائعة عند الطلاب، والخلط بين المعلومات والأفكار، أو سوء الفهم الذي يقع فيه الطلاب نتيجة لعدم إتقانهم للموضوع المختبر.

ونظراً لأن مفردة الاختيار من متعدد بمكوناتها الثلاث تعد وحدة متكاملة، لذلك يجب أن تتضمن المفردة فكرة أساسية، وأن يكون لها إطار مرجعي واحد للإجابة بحيث تقع المشتتات داخل هذا الإطار، بمعنى أنه لا ينبغي أن يكون أحد البدائل أو الإجابة الصحيحة من إطار مرجعي مختلف. فإذا كان الإطار المرجعي هو المكان، أو الزمان، أو الحدث فإن جميع المشتتات يجب أن تلتزم بأي من هذه الأطر، وليس جميعها. وإذا كانت المفردة من النوع الذي يتطلب اختيار أفضل الإجابات فإن جميع المشتتات ينبغي أن تكون صحيحة بدرجة ما، وعلى الطالب أن يفاضل بينها ويختار أفضلها.

ويُستخدم هذا النوع عندما يكون الهدف هو معرفة أفضل تعليل لموقفين، أو أفضل أسلوب لأداء عمل معين، أو أفضل تطبيق لقاعدة معينة، أو ما شابه ذلك. لذلك فإن طبيعة ونوع الهدف المراد قياسه هو الذي يحدد اختيار مفردة إما تتطلب إجابة صحيحة أو أفضل الإجابات.

ويمكن استخدام مفردات الاختيار من متعدد لقياس المستويات المعرفية العليا إلى جانب قياس تذكر بعض المعلومات وتعرف المعاني والحقائق النوعية. إذ يمكن بناء مفردات يختار فيها الطالب التعميمات المختلفة لتفسير ظاهرة معينة، وتحليل وتفسير بيانات معطاة، وتطبيق المبادئ أو مجموع مبادئ على مواقف مألوفة لديه، أو حل مشكلة أو مسألة تتميز بعنصر الجدة أو الحداثة، أو تقييم موقف أو قضية معينة، أو تركيب عناصر مختلفة من المعلومات بحيث تشكل إطاراً متكاملاً يتميز بالشمولية و التنظيم.

وينبغي مراعاة الشروط الآتية في كتابة مفردات الاختيار من متعدد:

قبل البدء في كتابة هذا النوع من المفردات يجب أن نراعي جيداً المستوى العرفي للهدف الإجرائي المطلوب قياسه وما يشير إليه الهدف، لأن هذا يساعد في بناء مفردات صادقة كما يزيد من ثبات درجات الاختبار.

ومن الأفضل البدء بكتابة كل مفردة بالقلم الرصاص على ورقة منفصلة أو بطاقة صغيرة، وأن يراجع شخص آخر يقوم بتدريس نفس المادة المفردات قبل إعداد الصورة النهائية للاختبار.

وفيما يلي بعض الاقتراحات التي تساعد في كتابة مفردات الاختيار من متعدد:

1. يجب أن تكون المشكلة المطلوب حلها أو الإجابة عليها كاملة في حد ذاتها بدون قراءة الإجابات المعطاة.
2. يجب أن تكون المشكلة خالية من الكلمات غير الضرورية، وتحتوي على معظم الكلمات التي قد تتكرر في أي من المشتتات أو الإجابات، وتجنب استخدام العبارات المنفية ما لم تتطلب ذلك نواتج التعلم الرئيسة.
3. يجب أن تتفق جميع الإجابات في صياغتها اللغوية مع العبارة الأساسية أو التمهيدية.
4. يجب أن تحتوي المفردات التي تقيس المستويات المعرفية العليا، مثل : الاستيعاب والتطبيق، على عنصر الجدة، ولكن في حدود ما تعلمه الطلاب.
5. يجب أن تكون جميع المشتتات متجانسة إلى حد ما حتى لا يستطيع الطلاب التوصل إلى الإجابة الصحيحة عن طريق استبعاد بعض المشتتات.

6. يجب أن نتجنب التباين الواضح في طول العبارات لكي لا يصبح طول العبارة بمثابة مؤشر للإجابة الصحيحة.

7. يجب أن يكون موضع الإجابة الصحيحة بين المشتتات المعطاة عشوائياً.

8. يجب عدم استخدام مشتتات خاصة مثل " ليست واحدة مما سبق"، أو " جميع ما سبق" إلا نادراً، وإذا استُخدم المشتت الأول ينبغي ألاّ يكون دائماً هو الإجابة الصحيحة.

9. يجب عدم استخدام مفردات الاختيار من متعدد إلا إذا كانت الأهداف المطلوب قياسها تتطلب ذلك.

10. يجب أن تكون الإجابة الصحيحة واضحة صحتها أو أفضليتها في علاقاتها بالإجابات الباقية. لكي نتأكد من أن الإجابات الأخرى ليست بنفس القدر من الصحة.

11. يجب اختيار المشتتات التي نرى أنه يمكن للطالب غير المستذكر اختيارها، على أن لا تكون غامضة أو غير معقولة.

12. يجب تجنب المشكلات التي تُبنى إجاباتها على آراء ذاتية.

13. يجب أن تكون المشتتات مستقلة بعضها عن البعض الآخر.

كيفية بناء مفردات الصواب أو الخطأ:

تُستخدم مفردات الصواب أو الخطأ True-False Items في قياس تذكر المفاهيم والمصطلحات، والقوانين والحقائق، والأحداث، وغير ذلك. كما تُستخدم في التمييز بين الحقائق والآراء، وبين الاستنتاجات الصحيحة والاستنتاجات الخطأ، والمقارنة بين الظواهر والأشياء. والحكم على صحة إجراء عمليات معينة، أو على شواهد وأدلة متنوعة، وغير ذلك من العمليات المعرفية البسيطة وتستخدم هذه المفردات بدرجة أكبر في المراحل التعليمية الأولى، حيث يضع التلميذ أمام كل عبارة علامة (✓) أو علامة (x)، أو (نعم، أو، لا) تبعاً لما تتطلبه المفردة.

ويتميز هذا النوع من المفردات بسهولة وموضوعية تصحيحه بواسطة أي شخص غير مؤهل لتدريس المجال الدراسي موضع الاختبار.

كما أن الاختبار الذي يتكون من مفردات الصواب أو الخطأ يمكن تطبيقه بسرعة، حيث إن

إجابة هذه المفردات تتطلب وقتاً قصيراً نسبياً إذا ما قورنت بغيرها من المفردات. والزمن اللازم لبناء هذه المفردات أقل من زمن بناء مفردات الاختيار من متعدد، حيث إنها لا تتطلب كتابة مشتتات لكل مفردة. ويشتمل الاختيار في هذه الحالة على عينة كبيرة من الحقائق يمكن أن يجيب عليها الطالب في زمن قصير نسبياً.

غير أن بعض علماء القياس والتقويم التربوي يقترح استخدام هذا النوع من المفردات كوسيلة تساعد المعلم على التأكد من معرفة الطلاب للحقائق المختلفة التي تقدم لهم أثناء العملية التعليمية، وليس لأغراض التقويم. وعلى الرغم من هذا الاقتراح، إلا أنه لا يزال كثيراً من المعلمين يستخدمون هذا النوع من المفردات في قياس الأهداف أو المستويات التعليمية التي تتطلب ذلك.

ومن العيوب الرئيسية لمفردات الصواب أو الخطأ أنها تشجع الطالب على تخمين الإجابة الصحيحة، إذ أن احتمال الإجابة الصحيحة عن طريق التخمين العشوائي 0.50 وعلى الرغم من أن هناك طرقاً لتصحيح الدرجة من أثر التخمين، مثل طرح عدد المفردات التي أجاب عليها الطالب إجابة خطا من عدد المفردات التي أجاب عليها إجابة صحيحة، إلا أنه لا يمكن إنكار أثر الصدفة والتخمين في التوصل إلى الإجابة الصحيحة. وكذلك لا تفيد هذه المفردات كثيراً في تعرف الفجوات في تحصيل الطلاب أو الأهداف التعليمية التي لم يحققونها. لذلك يصعب باستخدام هذه المفردات التمييز بين الطلاب ضعاف التحصيل، والطلاب أقوياء التحصيل.

ومن النقد الموجه إلى مفردات الصواب أو الخطأ أيضاً أنها لا تقيس إلا أدنى المستويات المعرفية، وهو تذكر واسترجاع الحقائق والمعلومات، ويندر استخدامها في قياس المستويات المعرفية العليا، وبذلك يصبح من غير الممكن بناء هذا النوع من المفردات لقياس مختلف الأهداف التعليمية المذكورة في جدول مواصفات الاختبار.

لذلك ينبغي على من يقوم بكتابة هذه المفردات أن يكون مدركاً لحدودها، ويختار من بين الأهداف أو المستويات التعليمية تلك التي يمكن قياسها باستخدام هذا النوع. ويمكن استخدام الأنواع الأخرى من المفردات عندما يستلزم الأمر ذلك لكي يتفق الاختبار مع جدول المواصفات مما يزيد من صدق الاختبار.

ومن الأفضل البدء بكتابة كل مفردة على بطاقة صغيرة حتى يسهل وضعها بالترتيب المطلوب ومراجعتها إذا استلزم الأمر ذلك.

وتتكون مفردات الصواب أو الخطأ من عبارة معينة يكون هناك اتفاق على صحتها أو عدم صحتها، وطريقة تدوين الإجابة عليها. والنوع الشائع منها يُطلب فيه من الطالب أن يضع علامة صواب أو خطأ فقط أمام المفردة، وأحياناً يُطلب منه أيضاً تحديد العنصر الخطأ، وتصحيح هذا الخطأ. ولكن هذا يؤدي إلى تدخل عنصر الذاتية في عملية تقدير الدرجات، إذ قد يعتبر المعلم الإجابة التي كتبها الطالب غير صحيحة على الرغم من إمكانية تبرير صحتها. وكذلك يستغرق الطالب وقتاً أطول للإجابة على مثل هذا النمط من مفردات الصواب أو الخطأ. وقد سبق أن ذكرنا أن إجابة عدد كبير من مفردات الصواب أو الخطأ المعتادة هي إحدى ميزات هذه المفردات.

ومن أمثلة مفردات الصواب أو الخطأ ما يلي:

التعليمات:

إقرأ كل عبارة مما يأتي، وضع دائرة حول الحرف (ص) إذا كانت العبارة صحيحة، وأخرى حول الحرف (خ) إذا كانت خطأ:

1. المستطيل أضلاعه متساوية. ص خ
2. اكتشف كولومبس قارة أستراليا. ص خ
3. الجذر التربيعي للعدد (81) يساوي (9). ص خ
4. الجزيئات ذات الشحنة الكهربائية السالبة تسمى نيوترونات. ص خ
5. المتوسط الحسابي هو أحد مقاييس النزعة المركزية. ص خ
6. كلمة (يفي) تعني (يزيد). ص خ

ومن الجدير بالذكر أنه يمكن ابتكار أنماطاً أخرى من مفردات الصواب أو الخطأ، كأن تُقدم أشكال أو رسوم، أو تمثيلات بيانية تتعلق بظواهر أو قضايا معينة، بحيث يليها عدد من مفردات الصواب أو الخطأ التي تقيس فهم الطالب لعلاقات أو تفسيرات معينة، أو الحكم على شواهد وأدلة مختلفة مستمدة من هذه المثيرات، وغير ذلك.

وينبغي مراعاة الشروط الآتية في كتابة مفردات الصواب أو الخطأ:

(1) يجب أن تشتمل كل مفردة على فكرة أساسية واحدة.

(2) يجب صياغة المفردة بحيث لا تحتمل أكثر من إجابة واحدة.

(3) يجب أن تكون العبارة قصيرة، وأسلوبها بسيط، ولغتها واضحة.

(4) ينبغي تجنب استخدام العبارات المنفية.

(5) ينبغي أن نذكر مصدر العبارات التي تعبر عن آراء شخصية.

(6) ينبغي تجنب استخدام ما يسمى "المحددات النوعية Specific Determiners"، والتي تكون بمثابة مؤشرات للحكم على صحة أو خطأ العبارات، مثل الكلمات: دائماً، أبداً، جميع، فقط، عادة، ربما، أحياناً.

(7) يجب أن تكون جميع العبارات متساوية في الطول تقريباً، لكي لا يصبح طول العبارة مؤشراً للإجابة الصحيحة.

(8) ينبغي تجنب نقل العبارات حرفياً من الكتب المدرسية.

مفردات المزاوجة:

تعد مفردات المزاوجة صورة خاصة من صور مفردات الاختيار من متعدد، فهي تتكون من سلسلة من مفردات الاختيار من متعدد، حيث تشترك مجموعة متجانسة من العبارات التمهيدية Premises في مجموعة البدائل أو الإجابات المناسبة لها Alternatives . وتوضع المجموعتان في قائمتين متوازيتين تشتمل إحداهما على العبارات التمهيدية أو الأساسية أو المثيرات، والأخرى على الإجابات الممكنة. ويقوم الطالب بمزاوجة مثيرات القائمة الأولى بالإجابات المناسبة لها في قائمة الإجابات وفقاً لتعليمات معطاة توضح كيفية إجراء المزاوجة، ويُكتب الرقم أو الحرف المناظر للمثير أمام الإجابة التي اختارها من القائمة الثانية.

ويمكن أن تتنوع المثيرات والإجابات في مفردات المزاوجة؛ فمثلاً يمكن المزاوجة بين المصطلحات والمفاهيم وتعاريفها، أو المفاهيم المختلفة ورموزها، أو المفردات اللغوية ومرادفاتها أو معانيها، أو القواعد والمبادئ وأمثلة توضحها، أو الأجهزة والأدوات واستخداماتها، أو

الشخصيات وأسمائها، أو المناطق الجغرافية وخصائصها، أو المدن وعواصمها، أو المعارك والفتوحات وقادتها، وغير ذلك.

ونظراً لهذا التنوع في المثيرات، فإن مفردات المزاوجة تعد من المفردات الاختبارية الأكثر فاعلية في قياس تذكر عدد كبير من الحقائق والمعارف في وقت قصير، وتشغل مساحة أقل بكثير مما تشغله مفردات الاختيار من متعدد التي تقيس هذا القدر من الحقائق والمعارف، إضافة إلي سرعة تصحيحها ودقته وموضوعيته .كما يمكن تطويع مفردات المزاوجة بحيث يمكن أن تقيس بعض المستويات المعرفية الأعلى من التذكر. إذ يمكن وضع مجموعة من المبادئ التي تُدرس في أي من المجالات الدراسية في القائمة الأولى (العمود الأيمن)، ومجموعة من المشكلات التي يمكن أن تُحل بواسطة هذه المبادئ في القائمة الثانية. (العمود الأيسر)، كما يمكن أستخدام الأشكال، والرسوم البيانية، والخرائط.

ويجب أن تبدأ مفردة المزاوجة بتعليمات واضحة تحدد الإطار المرجعي العام للاستجابة. وهذه التعليمات يمكن أن تتكون من عبارة واحدة أو عدة عبارات محددة تناسب المستوى النمائي للطلاب.

وفيما يلي مثالاً بسيطاً يوضح هذا النوع من المفردات:

تعليمات: فيما يلي مجموعة من الدول تقابلها مجموعة من الواحات. اختر الواحة التي تنتمي إلى الدولة المناسبة لها، ولا تختر أكثر من واحة لدولة واحدة، وضع الحرف المناظر للواحة بين قوسي الدولة المناسبة:

الدولة	الواحة
( ) السعودية	(أ) الفرافرة
( ) الجزائر	(ب) الهفوف
( ) مصر	(جـ) عين صالح
( ) ليبيا	(د) سيوة
	(هـ) الكفرة

وتنطبق معظم شروط كتابة مفردات الاختيار من متعدد، والصواب أو الخطأ على مفردات المزاوجة إضافة إلى الشروط الآتية:

1. يجب أن تكون الإجابات التي في القائمة الثانية مناسبة إلى حد ما لجميع العبارات التمهيدية أو المثيرات التي في القائمة الأولى، ومتفقة معها لغوياً، ومتجانسة بدرجة ما في المحتوى. وأن يكون لكل عبارة أو مثير إجابة واحدة. فكلما زاد تجانس هذه البدائل زادت صعوبة المفردة.
2. يجب أن يكون عدد الإجابات (البدائل) في القائمة الثانية أكبر من عدد المثيرات في القائمة الأولى.
3. يجب أن تحدد التعليمات إمكانية اختيار أكثر من بديل مع كل مثير؛ فاستخدام بعض البدائل أكثر من مرة يقلل من احتمال توصل الطالب إلى الإجابة الصحيحة عن طريق حذف البدائل التي يستخدمها في كل مرة عند إجابته.
4. يجب ألا يزيد عدد البدائل أو المثيرات في القائمة الأولى عن عشرة، لكي لا يستغرق الطالب وقتاً طويلاً في قراءتها كلما انتقل إلى البديل أو المثير التالي.
5. يجب ألا يكون هناك نمط سائد في ترتيب كل من المثيرات والإجابات.
6. يجب تجنب الأخطاء اللغوية، والجمل المعقدة عند صياغة عبارات المفردة.
7. يجب كتابة المفردة بأكملها في صفحة واحدة دون تجزئتها.
8. يجب وضع عناوين مناسبة موجزة لقائمة المثيرات وقائمة الإجابات.
9. يفضل أن تكون العبارات التمهيدية (المثيرات) أطول نسبياً من عبارات عناصر الإجابات.
10. ينبغي مراعاة أن مفردة المزاوجة تعد وحدة متكاملة يتطلب تغيير جزء منها إعادة النظر في المفردة ككل.

مفردات الإكمال، والإجابة القصيرة:

يتطلب هذا النوع من المفردات أن يكمل الطالب عبارة أو جملة ناقصة بكلمة من عنده، أو يجيب على سؤال إجابة قصيرة لا تتعدى بضع كلمات،أو فقرة محدودة، أو يربط بين ظاهرتين

معينتين. وتُستخدم هذ المفردات بكثرة في الكتب المدرسية لمراجعة الموضوعات المختلفة، كما تُستخدم في الاختبارات الصفية، وبخاصة في العلوم الفيزيائية والبيولوجية.

ولهذا النوع من الفردات بعض المزايا إذا استُخدم في قياس تذكر الحقائق والمعارف بدلاً من التعرف عليها كما هو الحال في المفردات التي سبق أن أوضحناها. ويمكن تطوير هذا النوع من المفردات، وبخاصة مفردات الإجابة القصيرة لقياس الفهم والتطبيق، مثل : صياغة فروض علمية، وتفسير رسوم وأشكال بيانية، وتطبيق معادلات رياضية، وإكمال وموازنة معادلات كيميائية، وحل مسائل حسابية، وتبرير قضايا، وتحديد أسباب أحداث معينة، وغير ذلك.

كما أن هذه المفردات تسمح بمعرفة العمليات التي اتبعها الطالب في التوصل إلى حل مسألة رياضية معينة، أو تفسير مشكلة محددة وصياغة ذلك بإيجاز، وهو ما لا يتأتى في المفردات التي تتطلب التعرف، حيث إنها تقيس الناتج النهائي فقط. وهذا يساعد المعلم في تشخيص الأخطاء العملياتية المختلفة.

وهذا النوع من المفردات أسهل نسبياً في كتابته من مفردات الاختيار من متعدد مثلاً، حيث إنها لا تحتاج إلى كتابة مشتتات يختار الطالب من بينها. غير أن الصعوبة الرئيسة في كتابة هذه المفردات هي التفكير في الإطار المرجعي الذي تعتمد عليه إجابة المفردة عند صياغة العبارة، بينما يتحدد هذا الإطار المرجعي بصورة جزئية عند كتابة مفردات الاختيار من متعدد، أو المزاوجة عن طريق استخدام المشتتات كما أن فرصة التخمين في إجابة مفردات الإكمال أو الإجابة القصيرة تكون قليلة مقارنة بمفردات الاختيار من متعدد، أو الصواب أو الخطأ. وإذا اعتُنى بكتابة مفردات الإكمال بحيث تستدعي إجابة واحدة، فإنه يمكن عندئذ عمل مفتاح تصحيح للإجابات مما يسهل عملية التصحيح ويجعلها أكثر موضوعية.

غير أنه يُعاب على هذا النوع من المفردات صعوبة استخدامه في قياس مهارات عقلية عليا، مثل: التحليل، والتركيب، والتقويم، التي تتطلب إجابات طويلة نسبياً. كما أن تكوين إطار مرجعي للإجابة على مفردات الإكمال يحتاج إلى قدر من الجهد عند كتابة هذه المفردات. ولتوضيح ذلك ننظر إلى المفردة الآتية:

" ما هي كرات الدم الحمراء؟"

فهذه المفردة يمكن الإجابة عليها بطرق مختلفة مثل " بروتينات"، " جزء من الدم"، " عدد خلايا" وهكذا. لذلك ربما يكون من الأفضل إعادة صياغة المفردة كالتالي:

" أي خلايا الدم تحمل الأكسيچين إلى خلايا الجسم البشري؟"

وكلمة " الدم" تعد بمثابة مؤشر يساعد في تحديد الإطار المرجعي لإجابة السؤال، أما إذا حذفنا هذه الكلمة، فإن المفردة تقيس في هذه الحالة قدرة الطالب على التمييز بين الخلايا ووظائفها، فتصبح المفردة كالتالي:

" أي خلايا الجسم البشري تحمل الأكسيچين إلى جميع الخلايا الأخرى؟"

لذلك ينبغي مراعاة الشروط الآتية في كتابة مفردات الإكمال والإجابة القصيرة:

1. يجب أن تحتوي المفردة على مؤشرات كافية فقط لتحديد الإطار المرجعي للإجابة، أما إذا كانت المفردة تحتوي على عدد كبير من المؤشرات، فإنها تصبح غاية في السهولة.
2. يجب أن تكون الصياغة اللغوية للمفردة صحيحة، وأن تُحذف جميع الكلمات غير الضرورية.
3. يجب أن يكون المكان الخالي في مفردة الإكمال في آخر العبارة بقدر الإمكان.
4. يجب عدم نقل العبارات مباشرة أي بنفس أسلوب صياغتها من الكتاب المدرسي، لأن هذا يشجع الطلاب على التذكر الآلي، كما يجعل العبارة أكثر غموضاً لانفصالها عن مضمونها الأصلي.
5. إذا كانت العبارة تتطلب عدة إكمالات يجب أن يكون كل منها مستقلاً عن الآخر، إلا إذا كان ترابط الإجابات يعد هدفاً نحتاج إلى قياسه.

فمثلاً يجب تجنب النوع التالي من العبارات:

" ثورة .... قد نشبت ....".

فهذه العبارة لا تحدد إطاراً مرجعياً للإجابة، ولكن الطلاب يستطيعون الإجابة عليها عن طريق تخمين قصد المعلم. ومع هذا إذا أخطأ في إكمال المكان الخالي الأول، فإن بقية الإجابات بالطبع ستكون خطأ.

لذلك يمكن إعادة صياغة هذه العبارة كالتالي مثلاً:

"نشبت ثورة القاهرة الأولى في عهد............ عام............ ."

6. يجب مراعاة أن تكون المسافات الخالية متساوية في الطول لكي لا يصبح طول العبارة مؤشراً للإجابة الصحيحة.

ولتيسير عملية التصحيح يُفضل إعداد ورقة إجابة منفصلة لمفردات الإكمال، بحيث تُكتب أرقام المفردات عليها. ويمكن إعداد مفتاح تصحيح يحتوي على الإجابات الصحيحة، وبذلك يمكن مضاهاة إجابة الطالب بالإجابة الصحيحة.

أسئلة المقال:

تعد أسئلة المقال أكثر أنواع المفردات استخداماً في الاختبارات الصفية التي يكتبها المعلم للطلاب، وهي تناسب قياس المستويات المعرفية العليا التي ربما لا تصلح كثير من المفردات الاختبارية الأخرى لقياسها بنفس القدر من الملاءمة والفاعلية. فسؤال المقال يتطلب أن يكتب الطالب إجابته بعد صياغتها بنفسه. غير أن كثيراً من المعلمين يستخدمون هذا النوع من الأسئلة في قياس تذكر الحقائق والمعارف في مختلف المواد الدراسية، وهذا لا يتمشى مع طبيعة هذه الأسئلة والأهداف أو المستويات التعليمية المناسبة لها التي ينبغي أن تقيسها. فالمفردات التي سبق أن أوضحناها تكون أكثر فاعلية وفعالية من أسئلة المقال في قياس التذكر والفهم. أما أسئلة المقال فينبغي أن يقتصر استخدامها في قياس مهارات التفكير المتعددة التي يجب أن يكتسبها الطلاب، ويعمل المعلمون على تنميتها لديهم من خلال المجالات الدراسية المختلفة، مثل:

- تنظيم وإعادة تنظيم حقائق معطاة.
- إجراء مقارنات، وتحديد أوجه الاتفاق والاختلاف بين ظواهر أو قضايا معينة.
- توضيح قواعد أو مبادئ بأمثلة مناسبة.
- إجراء تطبيقات للمبادئ تتميز بعنصر الجدة.
- التوصل إلى استنتاجات من بيانات أو معلومات معطاة.
- نقد قضايا معينة استناداً إلى محكات معطاة.
- اكتشاف مداخل جديدة لحل مشكلات أو مسائل معينة.

- تحليل الأفكار في الموضوعات المختلفة.
- التعبير اللغوي التحريري.
- كتابة التقارير والملخصات والخطابات وغيرها.
- اقتراح آراء وتبريرها بأدلة منطقية مناسبة.
- تقييم نتائج التجارب.

ومن هذا يتضح أن سؤال المقال يتيح للطالب الفرصة لإعمال عقله، وتنظيم أفكاره، والتعبير عنها، والمقارنة بين الموضوعات المختلفة، ونقدها أو تقييمها. وعلى الرغم من أنه يمكننا استخدام مفردات الاختيار من متعدد في قياس بعض هذه العمليات العقلية العليا، إلا أن كتابتها لهذا الغرض تحتاج إلى مهارة فنية كبيرة، ووقت طويل.

فمثلاً إذا كان معلم التاريخ يود قياس قدرة طلابه على استخدام المعلومات الاجتماعية والاقتصادية والسياسية بغرض التوصل إلى فروض علمية تتعلق بالتكوين الثقافي والسياسي للمجتمع، فإن سؤال المقال يناسب هذا الهدف بدرجة أفضل لكي يستطيع معرفة كيفية توصل الطالب إلى إجابته، وهنا لا يكون للتخمين تأثير في صحة الإجابة.

ويُنصح بعدم استخدام أسئلة المقال في المرحلة الابتدائية لأنها تحتاج في إجابتها إلى تعبير لغوي لا يتوافر لدى التلاميذ في هذه المرحلة المبكرة، ولكن يمكن استخدام هذه الأسئلة في المراحل الأعلى وبخاصة لقياس العمليات العقلية العليا في هذه المراحل، مع العناية بكتابتها وطريقة تصحيحها.

غير أنه يُعاب على أسئلة المقال انخفاض صدق محتوى الاختبار الذي يقتصر عليها، وذلك لأن إجابة هذه الأسئلة يحتاج إلى زمن أكبر من إجابة المفردات التي تتطلب إجابة مقيدة، مما يجعل الاختبار يقتصر عادة على عدد قليل من الأسئلة، مما يؤدي إلى انخفاض درجة تمثيلها للأهداف المهمة التي ينبغي تحقيقها من المجال الدراسي المعين.

كما أن تقدير درجات إجابات أسئلة المقال يستغرق وقتاً طويلاً نسبياً، وبخاصة إذا كان عدد الطلاب كبيراً بعكس تصحيح المفردات التي سبق أن أوضحناها. ويمكن تلافي ذلك إلى حد ما إذا تمت صياغة السؤال بحيث يتطلب إجابة قصيرة نسبياً. ونظراً لصعوبة موضوعية

تقدير هذه الدرجات، فإن هذا التقدير ربما يتأثر بعوامل ذاتية، مثل تأثير الهالة Hallo Effect وهذا يحدث عندما يؤثر الانطباع العام للمعلم عن ورقة الإجابة على تقديره لدرجات في هذه الورقة، وسوف أوضح بعد قليل كيفية التغلب على ذلك.

كذلك ربما تلعب بعض العوامل الخارجة عن موضوع السؤال في تقدير إجابة السؤال مثل: استخدام القلم الرصاص بدلاً من القلم الحبر في الإجابة، أو طريقة الكتابة، أو الأخطاء اللغوية. لذلك يجب على المعلم أن يبذل قصارى جهده لاستبعاد هذه العوامل، ويركز على تقدير عناصر الإجابة المرتبطة بالهدف أو المستوى التعليمي standard المراد قياسه.

وبناء أسئلة مقال جيدة وخالية من الغموض لا يعد أمراً يسيراً، وإنما يتطلب مراعاة شروط متعددة، من أهمها ما يلي:

1. يجب استخدام أسئلة المقال لقياس المستويات المعرفية العليا فقط، إذ يمكن قياس المستويات الأدنى بدرجة أكثر فعالية باستخدام أي من أنواع المفردات التي سبق توضيحها.
2. يجب صياغة السؤال بحيث تكون المشكلة واضحة ومحددة للطالب. ونظراً لأن أسئلة المقال يجب أن نستخدمها لقياس المستويات المعرفية العليا، فيجب أن نتجنب استخدام كلمات مثل: من، ماذا، متى، أين، اذكر، لأن المفردات التي تقيس مستوى التذكر تبدأ إعادة بمثل هذه الكلمات. أما المفردات التي تقيس المستويات المعرفية العليا فعادة تبدأ بكلمات مثل: لماذا، فسر، قارن، اربط، بيِّن أوجه الاختلاف، حلل، انقد، قوِّم، وما إلى ذلك.
3. يجب عدم وضع أسئلة اختيارية إلا إذا كان الناتج التعليمي يستدعي ذلك، لأن إعطاء فرصة الاختيار يجعل من الصعب مقارنة الطلاب بعضهم بالبعض الآخر، كما يشجعهم على ترك أجزاء من المادة الدراسية بدون مذاكرة، وهذا بالطبع يقلل من مدى تمثيل الاختبار لأهداف المادة الدراسية المطلوب قياسها، وبالتالي يقلل من صدق محتوى الاختبار.
4. يجب تحديد الوقت المخصص لإجابة كل سؤال على أن يكون هذا الوقت كافياً للتفكير والإجابة.

5. يجب أن تكون تعليمات الإجابة واضحة وضوحاً تاماً لجميع الطلاب، ونعني بهذا الزمن اللازم لإجابة كل سؤال، والمعلومات المطلوب كتابتها في الإجابة ونوع الإجابة، حتى لا يستخدم الطالب أسلوب التخمين لمعرفة قصد المعلم من السؤال.

ويمكن التغلب على هذه المشكلة بأن يتضمن كل سؤال تعليمات واضحة للإجابة وكذلك الأوزان النسبية لكل سؤال.

6. يجب عمل نموذج إجابة واضح لكل سؤال. فبعد كتابة السؤال ومراجعته، يجب على المعلم أن يكتب نموذج لما يعتبره أفضل إجابة للسؤال، وهذا يساعده على كشف نواحي الغموض وعدم الاتساق في صياغة السؤال، كما أنه يزيد من ثبات درجات الاختبار.

طرق تصحيح أسئلة المقال:

يمكن الاسترشاد بالقواعد التالية عند تصحيح أسئلة المقال لكي تقلل من العوامل الذاتية عند تقدير درجات الإجابة:

1. يجب تقدير درجات أسئلة المقال في ضوء النواتج أو الأهداف التعليمية المطلوب قياسها. فاختبار المقال شأنه كشأن الاختبار الموضوعي يُستخدم لقياس تحصيل الطالب للنواتج التعليمية المطلوبة. فإذا كان السؤال يقيس "القدرة على تفسير علاقة السبب بالنتيجة" مثلاً، فإننا يجب أن نراعي مدى كفاية تفسير الطالب لعلاقة السبب بالنتيجة عند تقدير درجة إجابته. ونستبعد العوامل الأخرى مثل: أسلوب الطالب في الكتابة، أو المعلومات التي لا ترتبط بالموضوع وإن كانت صحيحة، أو الأخطاء اللغوية، وغير ذلك.

2. يمكن تقدير درجات أسئلة المقال التي تحتاج إلى إجابة قصيرة نسبياً باستخدام طريقة النقط Points.

وهنا نكتب نموذج إجابة لكل سؤال، ونحدد عدد النقط التي تُعطى لكل جزء من الإجابة، مع مراعاة الأهمية النسبية للنواتج التعليمية التي تقيسها الأسئلة.

3. يمكن تقدير درجات الإجابة على الأسئلة التي تحتاج إلى إجابة طويلة باستخدام طريقة التصنيف Sorting.

وهنا نقرأ أوراق الإجابة ثم نصنفها في مجموعات مختلفة، مثل: مجموعة ممتازة، ومجموعة جيدة جداً، ومجموعة جيدة، ومجموعة متوسطة، ومجموعة غير جيدة.

ويجب تحديد عدد هذه المجموعات قبل بدأ عملية التقدير للدرجات.

وإذا لم نستطع تحديد المجموعة التي تنتمي إليها أحد أوراق الإجابة، فإننا يجب إعادة قراءتها بعناية حتى يتسنى لنا تصنيفها في إحدى المجموعات. ويجب التأكد من أن أفقر الإجابات في المجموعة الممتازة تفوق أفضل الإجابات في المجموعة غير الجيدة.

ومن الأفضل بعد الانتهاء من عملية التصنيف المبدئي أن نعيد تصنيف أوراق الإجابة في كل مجموعة للتأكد من تجانسها. والهدف من إعادة التصنيف أن نجعل الفروق بين المجموعات أكبر ما يمكن، والفروق داخل كل مجموعة أقل ما يمكن. ومن ثم نضع تقديراً لكل مجموعة من هذه المجموعات مثل: ممتاز، جيد جداً، جيد، مقبول، راسب، وهكذا، على حسب عدد المجموعات التي نحددها قبل بدء عملية تقدير الدرجات. وسواء استخدمنا طريقة النقط أو طريقة التصنيف، فإنه يمكن زيادة موضوعية التصحيح إذا استبعدنا أسماء الطلاب من على أوراق الإجابة، وقررنا كيفية تقدير الإجابات الصحيحة ولكن غير المتعلقة بموضوع السؤال، وأيضاً الأخطاء اللغوية. كما أنه من المفيد أن نكتب بعض التعليقات على الإجابات حتى يتسنى للطلاب معرفة أخطائهم، مما ييسر لهم عملية التعلم.

(ثامناً) تقييم المفردات والمهام الاختبارية:

يتطلب الاختبار أو الأداة المستخدمة في قياس تحصيل الطالب ونتاجاته مراجعة وتدقيقاً بعد تطبيقه. إذ يمكن أن يتبين من النتائج قصور بعض المفردات، حيث تُترك بعض المفردات دون إجابة من ذوي التحصيل المرتفع، في حين يجيب عليها ذوي التحصيل المنخفض إجابة صحيحة. لذلك ينبغي إجراء تحليل المفردات Item Analysis الذي سبق أن أوضحناه في الفصل الرابع، للتعرف على مثل هذه المفردات وغيرها التي تكون غامضة، أو مربكة، أو حدث خطأ في تحديد إجابتها الصحيحة في مفتاح التصحيح، وغير ذلك. وهذا يفيد المعلم في إجراء التعديلات المناسبة على هذه المفردات، وتجميعها في قوائم تشتمل على المفردات، وقيمة معامل صعوبة كل منها ومعامل تمييزها. وتوجد برامج جاهزة متعددة يمكن استخدامها في إجراء هذا التحليل بسرعة ودقة، مثل برنامج: MicroCAT، وبرنامج Iteman، أعدتهما مؤسسة نظم التقييم (ASC) الأمريكية.

وبذلك يمكن أن يستخدم المعلم هذه المفردات مرات متعددة في أعوام متتالية مما يوفر جهده ووقعته وتكريسهما لعملية تعليم طلابه.

الاختبارات التحصيلية المقننة Standardized Achievement Tests:

تُستخدم الاختبارات التحصيلية المقننة في قياس درجة تعلم الطالب في مجالات مشتركة للمناهج المدرسية. فالاختبارات التي يكتبها المعلم لطلابه تقتصر على ما يقوم بتدريسه وما يناسب حاجات طلابه، أما الاختبارات التحصيلية المقننة فإنه يمكن تطبيقها على طلاب فرقة دراسية معينة في مختلف المدارس في ظروف مقننة منتظمة من حيث التطبيق، والتصحيح. وهذا يجعل من الممكن مقارنة مستوى تحصيل طلاب مجموعة معينة بمعيار الجماعة Reference Group. فكما سبق أن ذكرنا في الفصل الرابع أن التقنين Standardization هو الاسلوب المنتظم الذي نتبعه عند تطبيق الاختبار، وتصحيحه لكي تكون المقارنة عادلة، حيث تتوحد الشروط والظروف لجميع الطلاب الذين يطبق عليهم الاختبار. وهذا المطلب يعد شرطاً أساسياً لجميع المشاهدات العلمية.

لذلك يتضمن الاختبار تعليمات تفصيلية واضحة تُراعى عند تطبيق الاختبار، لضمان انتظام الإجراءات والظروف التي يطبق فيها الاختبار. وتعد صياغة هذه التعليمات جزءاً من عملية تقنين هذا النوع من الاختبارات.

وكذلك يجب أن تتضمن عملية التقنين جميع المواد الاختبارية المستخدمة، وزمن الإجابة على الاختبار، وامثلة توضيحية، وغير ذلك من تفاصيل الموقف الاختباري الذي يمكن أن يؤثر في أداء الطلاب في الاختبار. وهذا التقنين يوفر للاختبار نوعاً من الموضوعية. كما يضمن عدم تأثر الدرجات باختلاف المصححين.

ونحن نحتاج في مجال التربية إلى معلومات نحصل عليها من عينة متسعة مقننة من أداء الطلاب. ومثال ذلك: قياس النمو التعليمي من عام إلى آخر، ومقارنة النتائج أو طرائق التدريس عبر المدارس، ومدى انتقال أثر التعلم، وتصنيف الطلاب. ويصعب الحصول على معلومات كافية في كل من هذه المواقف باستخدام الاختبارات الصفية التي يكتبها المعلم لطلابه. ومن هنا تبرز الحاجة إلى بناء الاختبارات التحصيلية المقننة التي يقوم بإعدادها خبراء متميزين في القياس النفسي والتربوي، وتقوم بنشر هذه الاختبارات مؤسسات نشر تجارية Test Publishers، وتشارك في إعدادها.

خصائص الاختبارات التحصيلية المقننة:

(1) تشتمل هذه الاختبارات على مهام يشترك جميع الطلاب في تعلمها في جميع المدارس، وتكون تعليماتها محددة لكي يبدأ الطلاب الإجابة عليها بفهم مشترك. فجميع الطلاب يجيبون على مهام تتطلب تذكر الحقائق، وحل المشكلات، والاستدلال مثلاً. وهذه المهام المشتركة تعد بمثابة الميزان الذي يجعل المقارنة بين المجموعات المختلفة من الطلاب ممكنة.

(2) تشتمل هذه الاختبارات على معايير Norms لتفسير الدرجات المستمدة منها، وذلك لتحديد مركز الطالب بالنسبة لأقرانه فيما يقيسه الاختبار. وعادة يتم تحويل الدرجات الخام إلى نوع من الموازين المعيارية، مثل: الرتب المئينية، والتساعيات المعيارية، كما أوضحنا في الفصل الرابع. وهذا يسمح بترتيب الطالب بين أفراد مجموعة كبيرة من الطلاب الآخرين، مما ييسر التعميم فيما يتعلق بالدرجات الناتجة.

(3) يشتمل معظم هذه الاختبارات على صيغتين متكافئتين أو اكثر، وتحتوي كل من هذه الصيغ على مجموعات متمايزة من المفردات، ولكنها تعد بمثابة عينة ممثلة لنفس النطاق السلوكي. ويكون متوسط درجات مجموعة من الطلاب في إحدى هذه الصيغ مساوياً المتوسط في صيغة اخرى، وكذلك يتساوى الانحراف المعياري لدرجاتهم في الصيغتين. وهذا يساعد في اختبار الطلاب قبل عملية التعليم وعقب الانتهاء منها دون أن تتأثر درجاتهم في المرتين بعوامل الذاكرة، وذلك لتعرف التغير الذي يحدث في هذه الدرجات.

(4) يُرفق مع كل من هذه الاختبارات دليل للاختبار Test Manual يُسترشد به في كيفية تطبيق الاختبار وشروط ذلك، وكيفية تفسير درجاته. وهذا يساعد في توحيد الإجراءات والتفسير.

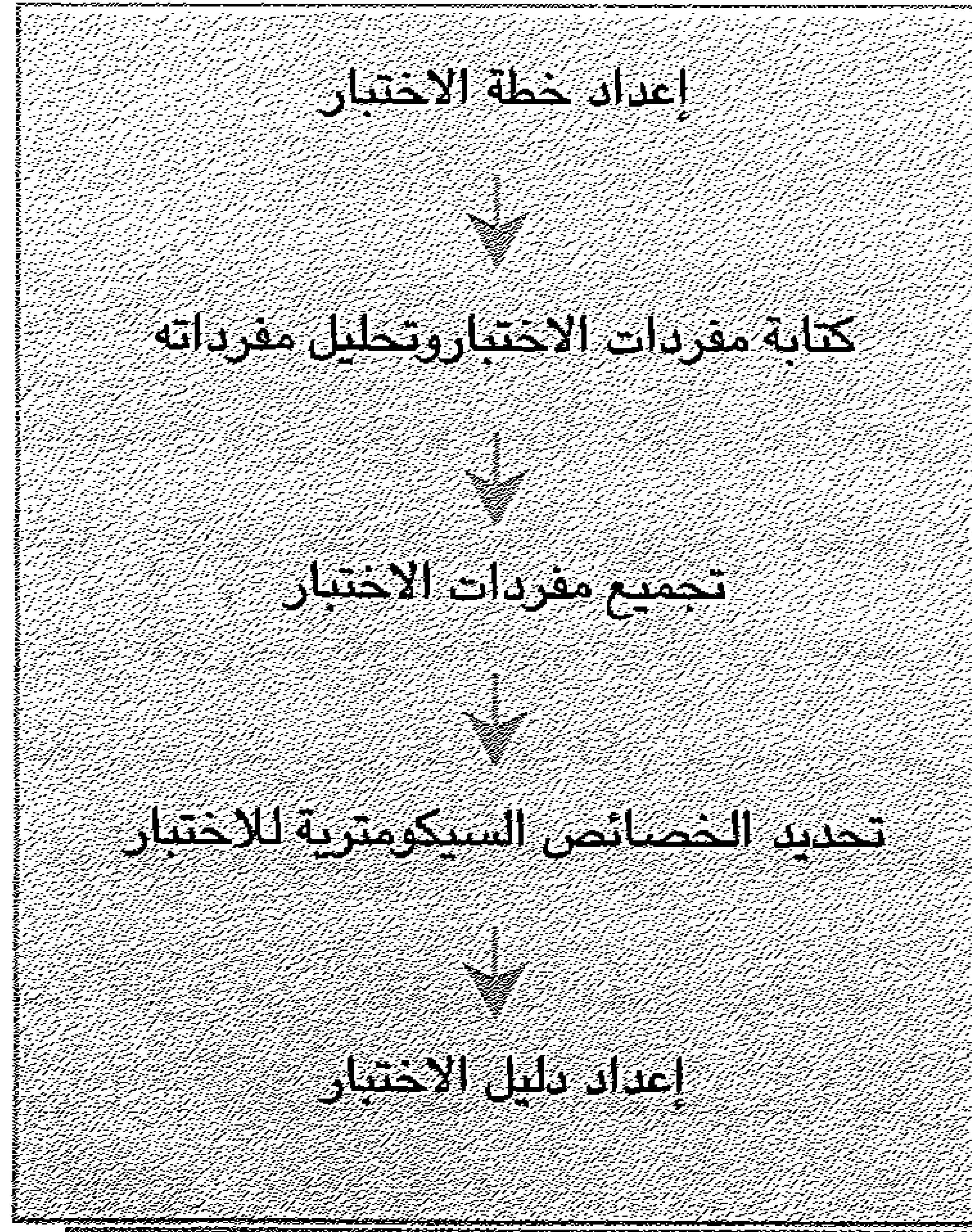
(5) تعد هذه الاختبارات من نوع الاختبارات مرجعية المعيار (NRTs) التي تهتم بمقارنة أداء الطالب بمحك أداء مطلق.

غير أن بعض الاختبارات التحصيلية المقننة يتم تصميمها وبناؤها بحيث تسمح بتفسير الدرجات استناداً إلى معايير، وكذلك إلى محك نجاح أو مستوى أداء محدد مسبقاً في ضوء الأهداف التعليمية المرجوة. وهناك بعض الاختبارات الأجنبية المتوافرة التي تفي بذلك في مجالات القراءة والحساب.

بناء الاختبارات التحصيلية المقننة:

يقوم بتصميم وبناء الاختبارات التحصيلية المقننة مؤسسات متخصصة، وعادة تكون مؤسسات نشر الاختبارات التربوية والنفسية في الولايات المتحدة الأمريكية. وتكون عملية بناء هذه الاختبارات تحت إشراف وتوجيه مجموعة استشاريين من خبراء القياس والتقويم واختصاصيي المناهج. وهؤلاء يضعون الخطوط العريضة للاختبار في ضوء الغرض من استخدامه، أي ما إذا كان مسحياً أو تشخيصاً، وكذلك يقومون بتحديد محتواه والمهارات المزمع قياسها، وأوزانها النسبية في ضوء المحتوى، والمستوى العمرى الملائم لمن سيطبق عليهم الاختبار، وطول الاختبار، والعلاقة بينه وبين غيره من الاختبارات التى ستتكون منها البطارية، أو الاختبارات التى سبق إعدادها ونشرها، ونوع المفردات التى ستُستخدم، وعدد الاختبارات الفرعية التى سوف تشتمل عليها البطارية، حيث سيكون لكل اختبار فرعى درجة مستقلة.

وعلى الرغم من أن خطوات بناء الاختبار المقنن ربما تختلف قليلاً بحسب الغرض منه وطبيعته، إلا أن هناك تسلسلاً معيناً ينبغي مراعاته يوضحه الشكل التخطيطي (5 - 5) التالى:



شكل (5 - 5) يوضح خطوات بناء الاختبارات التحصيلية المقننة

وفيما يلي توضيحاً لكل من هذه الخطوات:

إعداد خطة الاختبار :

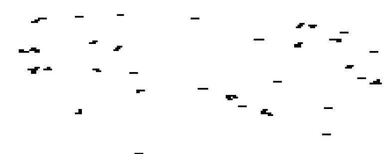
الفرق الرئيس بين الاختبارات التحصيلية المقننة والاختبارات التى يكتبها المعلم لطلابه هو أن الاختبارات المقننة يقوم بإعداد خطتها مجموعة من الخبراء، وتعكس هذه الخطة آراء ووجهات نظر هذه المجموعة، مما يساعد في تقليل عوامل التحيز. ويكون عادة محتوى الاختبار عريضاً وأكثر شمولاً، كما يكون عادة جزءاً من بطارية اختبارات تناسب فرقة دراسية معينة، ويكون الاختبار مناسباً للبطارية التى هو جزء منها.

وفيما عدا ذلك فإن خطة بناء الاختبار التحصيلى المقنن تماثل خطة بناء الاختبار الذى يكتبه المعلم. ويبدأ إعداد الاختبار بالفحص الدقيق للمادة الدراسية، والكتب المقررة لتحديد الموضوعات التى سيتناولها الاختبار، وربما يتطلب ذلك استشارة مجموعة من اختصاصيي المادة الدراسية والاستنارة بمقترحاتهم وتوصياتهم في هذا الشأن. وتُحدد المهارات الرئيسة والفرعية المتعلقة بهذا المحتوى، ويتم إعداد جداول المواصفات التى يُعين فيها المحتوى والمهارات المطلوبة، والأوزان النسبية لكل منها.

كتابة المفردات:

يقوم اختصاصيو المواد الدراسية – بالاستعانة بخبراء القياس – بكتابة مفردات الاختبار المناسبة طبقاً للخطة الموضوعة، وتُراجع بواسطة مجموعة أخرى من الخبراء، ومن ثم تُعاد صياغة المفردات ومراجعتها، وتتكرر هذه العملية حتى يتم الاتفاق على صلاحية هذه المفردات. وعادة يُبنى عدد من المفردات أكبر مما هو مطلوب، وذلك لأن بعض المفردات ربما يُستبعد لعدم صلاحيته. ويعتمد عدد المفردات التي يتم كتابتها لكل موضوع على الأوزان المحددة في جدول مواصفات الاختبار الذي تم إعداده، وعلى عدد الصيغ المتكافئة للاختبار.

وعادة يُستخدم في هذا النوع من الاختبارات مفردات اختيار من متعدد، وذلك لأنها أكثر تمثيلاً للمحتوى والمهارات المراد قياسها، كما أن تصحيحها يكون أكثر سهولة وموضوعية. غير أن استخدام مفردات أخرى مختلفة عن المفردات الموضوعية يمثل مشكلة تتعلق بتصحيح هذه المفردات. وعلى الرغم من البحوث التي تحاول جعل تصحيح أسئلة المقال موضوعياً، بل



واستخدام الحاسوب في تقدير درجات هذا النوع من الأسئلة، إلا أن هذه المحاولات لا تزال مبكرة. لذلك اتجه خبراء القياس والتقويم في الآونة الأخيرة الى استخدام أساليب التقويم البديل التي تختلف عن اساليب بناء الاختبارات المقننة، وذلك لقياس المهارات التي لا تستطيع المفردات الموضوعية قياسها.

التطبيق المبدئي للاختبار، وتحليل مفرداته:

بعد وضع خطة الاختبار وكتابة مفردات وتدقيقها يجب أن يطبق الاختبار تطبيقاً مبدئياً على عينة أو عينات من الطلاب تماثل المجموعة الفعلية المستهدفة التي سيُطبق عليها الاختبار في صورته النهائية. وهذه الخطوة تعد من الخطوات الضرورية، لأنه سيتم بناءً على نتائج هذا التطبيق تحليل المفردات Item Analysis ومراجعتها، وحذف المفردات غير المميزة، أو الغامضة، أو التي مستوى صعوبتها مرتفع، وغير ذلك، ومراجعة وتعديل هذه المفردات بما يضمن اشتمال الاختبار على أفضل المفردات.

تجميع مفردات الاختبار:

عند تجميع مفردات الاختبار في صورته النهائية يجب أن تمثل المفردات التي يشتمل عليها الاختبار جميع أقسام المحتوى والمهارات طبقاً للخطة الموضوعة. كما يجب أن يكون عدد المفردات مناسباً للزمن المحدد للإجابة عليها، وأن يكون مستوى صعوبتها مناسباً. ومثل هذا التوازن يعد من العمليات الدقيقة التي يجب أن تعتمد على البيانات الإحصائية، والأحكام المستنيرة.

تحديد الخصائص السيكومترية للاختبار:

إن تحديد الخصائص السيكومترية للاختبار المقنن يعد عملية مستمرة؛ فعلى الرغم من أن بعض خصائص مفردات الاختبار، مثل: تمييز المفردة، وصعوبتها، وزمن الإجابة عليها، يمكن تقديره من خلال نتائج التطبيق المبدئي للاختبار، إلا أن إجابات الطلاب عند التطبيق النهائي للاختبار ربما تختلف عن إجابة الطلاب الذين طبق عليهم الاختبار في التجربة الاستطلاعية المبدئية. لذلك فإنه يجب تحليل الصورة الكاملة للاختبار لمعرفة مدى اتساق الدرجات، وصدق

الاختبار، وبخاصة صدق محتواه. وينبغي اختيار معامل الثبات المناسب لطبيعة الاختبار، والغرض منه، والمجموعة المستهدفة. ونظراً لأن تفسير درجات الاختبار ينبغي أن يستند إلى إطار مرجعي معين، فإن هذا يتطلب إعداد معايير الاختبار Test Norms التي اوضحناها في الفصل الرابع، ويكون التفسير في هذه الحالة مرجعي المعيار Norm - Referenced، أي يُقارن أداء الطالب في الاختبار بأداء الجماعة أو الجماعات المرجعية التي تمثل الفرق الدراسية المختلفة، والمناطق الجغرافية، وحجم المدارس التي سيتم انتقاء عينات الطلاب منها.

غير أن هذا التفسير لا يقدم دليلاً على درجة كفاءة الطالب في المادة الدراسية، مما قد يستلزم تحديد محكات أو مستويات يُستند اليها في تحديد مقدار كفاءة أو مهارة الطالب في المادة الدراسية، وهذا يسمح بأن يكون التفسير مرجعي المحك Criterion - Referenced .

إعداد دليل الاختبار:

لا ينتهي إعداد الاختبار التحصيلي المقنن ببنائه واعداد جداول معاييره، وإنما ينبغي إعداد دليل أو أدلة الاختبار Test Manuals والمواد المساعدة، لمن يستخدمون الاختبار، بحيث تزودهم بمعلومات كافية لتحديد ما إذا كان الاختبار مناسباً لاغراضهم، وتُمكنهم من تفسير الدرجات تفسيراً صحيحاً. ويتبع معظم ناشرى الاختبارات في إعداد هذه الأدلة العناصر والمستويات التى حددتها الجمعية الأمريكية لعلم النفس (APA) في مطبوعها بعنوان "مستويات الاختبارات النفسية والتربوية وأدلتها".

وهذه المجموعة من المستويات ينبغى أن تكون مألوفة لدى من يود استخدام الاختبارات المقننة. وسوف أوضح ذلك في الفصل التاسع.

فجميع هذه الاختبارات ينبغي أن يصاحبها أدلة الاختبار التي تتضمن معلومات تتعلق بتفسير درجات الاختبار، وثباتها، وصدق الاختبار، وتعليمات عن تطبيق الاختبار وتصحيحه، والمعايير المناسبة للاختبار في صورة جداول تشتمل على الدرجات الخام الممكنة في الاختبار، والدرجات المحولة المناظرة لها التي تكون عادة مئينيات، أو تساعيات معيارية. وينبغي أن تكون هذه المعايير حديثة، وجميع المعلومات مكتملة.

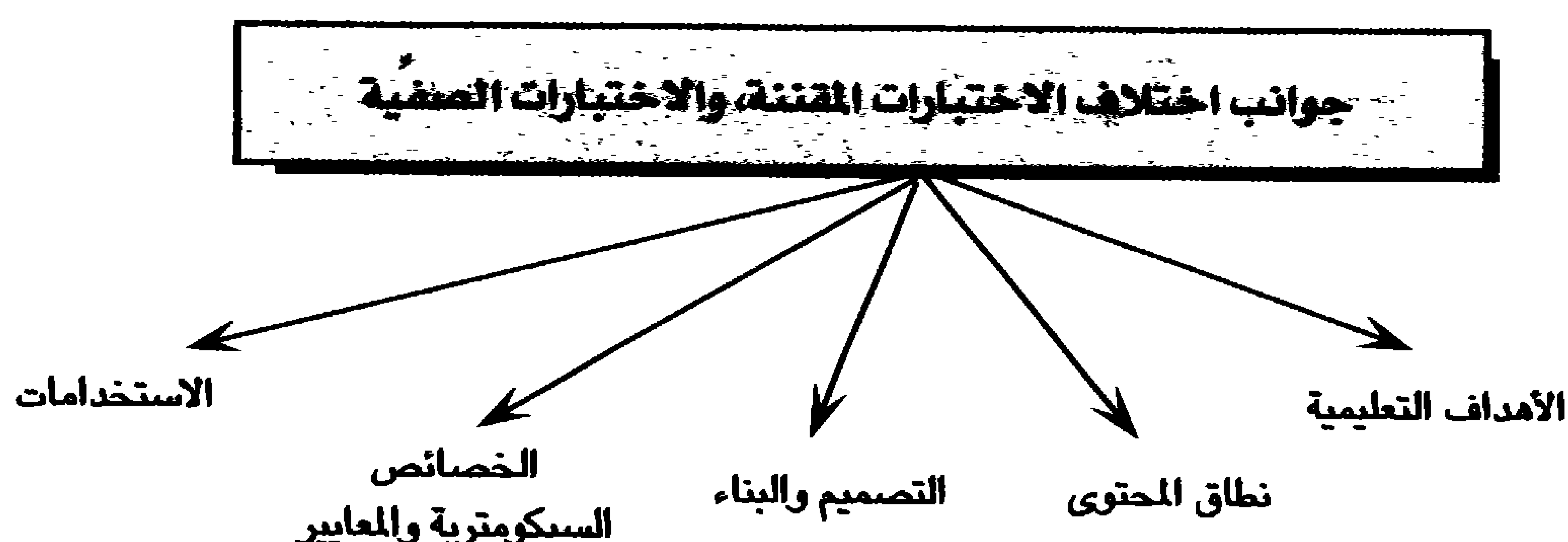
وعندما يود ناشر الاختبار إعداد دليلين أو أكثر للاختبار نفسه، أحدها يناسب المعلمين والموجهين وغيرهم من المربين، فإن الدليل يمكن أن يتضمن معلومات يمكن فهمها بسهولة دون التركيز على الجوانب الإحصائية المعقدة التى تحتاج إلى أخصائيين مدربين على تفسير البيانات السيكومترية.

أما الدليل الآخر فيكون دليلاً فنياً يتضمن بيانات ومعلومات إحصائية تتعلق بالاختبار، ومعاييره، وخصائصه السيكومترية، وذلك للمتخصصين الذي يودون فحص الاختبار فحصاً متعمقاً. وبذلك يكون الاختبار التحصيلى المقنن صالحاً للنشر.

ومن هذا يتضح أن بناء هذا النوع من الاختبارات يعد عملاً فنياً متخصصاً، ويتطلب قدراً كبيراً من الجهد والوقت والكلفة، ولذلك يوكل هذا الأمر إلى مؤسسات متخصصة لديها الخبرات الفنية المتميزة في القياس والتقويم، والموارد المالية اللازمة لذلك.

الفروق بين الاختبارات التحصيلية المقننة، والاختبارات الصفيَة:

يمكن استخدام الاختبارات التحصيلية المقننة إلى جانب الاختبارات الصفية في تقويم الطلاب، حيث إن كلاً منهما يُزوِّد المعلم بمعلومات مختلفة عن تحصيل طلابه، على الرغم من أن كليهما يقيس معارفهم ومهاراتهم. غير أن هناك فروقاً أساسية بين نوعى الاختبارات في عدد من الجوانب الموضحة بالشكل التخطيطي (6 - 5) التالى:



شكل (6 - 5) يوضح جوانب اختلاف الاختبارات التحصيلية المقننة، والاختبارات الصفية

الأهداف التعليمية : يستند المعلم في بنائه للاختبارات الصفِّية التى يكتبها لطلابه إلى أهداف تعليمية لوحدات دراسية يقوم بتدريسها في الصف المدرسى. وهذه الأهداف ربما تختلف من صف إلى آخر بما يناسب طلاب الصف المعين. وهذه تُعد من ميزات هذه الاختبارات في تحديد الأهداف التى حققها معظم الطلاب، والأهداف التى تتطلب مزيداً من الجهد لتحقيقها. أما الاختبارات التحصيلية المقننة فتُصمم لتقييم أهداف تربوية عريضة ومتسعة تناسب فرقة دراسية معينة.

نطاق المحتوى : تتعلق الاختبارات الصفِّية بمحتوى وحدة دراسية معينة أو موضوعات محددة، أما الاختبارات التحصيلية المقننة فإنها تتناول مجالات محتوى أكثر اتساعاً، لأنها ربما تُستخدم في تقييم تحصيل الطلاب في أكثر من عام دراسى واحد، وكذلك إذا كانت تشتمل على عدة اختبارات فرعية ضمن بطارية من الاختبارات. لذلك ينبغى معرفة الغرض من الاختبار، ومدى ملاءمته لقياس تحقُّق أهداف تعليمية معينة.

التصميم والبناء : يقوم بإعداد الاختبارات التحصيلية المقننة نخبة من خبراء القياس والتقويم واختصاصيي المواد الدراسية المختلفة، ويستغرق بناؤها وقتاً طويلاً، وجهدا وافراً، وكلفة عالية. ويمر بناء مفرداتها بمراحل متعددة للتحقق من توافر الشروط الفنية في كتابتها وصياغتها وتدقيقها، وكذلك تجريبها، وتحليل مفرداتها باستخدام الأساليب الإحصائية لتحديد معامل صعوبتها وتمييزها بين الطلاب،ومراجعتها أو تعديلها في ضوء ذلك. أما الاختبارات الصفِّية فإن المعلم يقوم بكتابة مفرداتها على الرغم من أنه ربما لا يكون لديه خبرة كافية أو تدرياً في فنيات بناء مختلف أنواع المفردات، ولا يقوم في كثير من الأحيان بتجريب المفردات للتحقق من جودتها باستخدام أساليب تحليل المفردات.

الخصائص السيكومترية والمعايير : تتميز الاختبارات التحصيلية المقننة بتحديد خصائصها السيكومترية، مثل الثبات، والصدق، ويزداد ثبات درجاتها في كثير من الأحيان عن (0.90)، مما يجعل هناك ثقة في اتساق الدرجات.

أما الاختبارات الصفِّية فإن المعلم نادراً ما يقوم بالتحقق من ثبات درجاتها، وإذا تم ذلك فإن معامل الثبات يكون عادة منخفضاً، وربما يرجع ذلك إلى تذبذب درجات الاختبار، وقلة عدد مفرداته.

وتتميز الاختبارات المقننة كذلك بأنها تشتمل على معايير من نوع معين؛ فبعضها يشتمل على معايير قومية يمكن في ضوئها مقارنة أداء طالب، أو صف، أو فرقة، أو مدرسة معينة بالتقدم الأكاديمي للطلاب على المستوى القومي.

وبعض آخر يشتمل على معايير محلية أو على مستوى مناطق معينة. لذلك فإن هذه الاختبارات تكون مرجعية المعيار. أما الاختبارات الصفِّية فلا يتوافر لها هذه المعايير، حيث إنها تكون عادة مرجعية المحك.

الاستخدامات : على الرغم من أن الغرض من الاختبارات المقننة، والاختبارات الصفِّية هو تقييم معارف الطلاب ومهاراتهم في وقت معين، إلا أنه نظراً لاختلافهما في المنظور والمحتوى، والبيانات المتعلقة بالصدق والثبات والمعايير، فإنهما يختلفان في الاستخدامات. فالاختبارات الصفِّية تُستخدَم في تقييم مدى تحقق أهداف تعليمية محددة بدرجة أفضل من الاختبارات التحصيلية المقننة، أما الاختبارات التحصيلية المقننة فإنه يمكن الاستناد إليها في اتخاذ قرارات تتعلق بالتوجيه والإرشاد التربوى والمهنى للطلاب، وانتقائهم وتسكينهم ورفعهم، وتقويم المناهج وأداء المعلم، وإجراء البحوث، وكذلك المقارنة بين الطلاب في المدارس المختلفة، وغير ذلك.

ويمكن تلخيص الفروق بين الاختبارات التحصيلية المقننة، والاختبارات الصفِّية في جدول (5 - 3) التالي:

جدول (5 - 3) تلخيص الفروق الأساسية بين الاختبارات التحصيلية المقننة، والاختبارات الصفِّية

	الاختبارات الصفِّية	الاختبارات المقننة
نواتج التعلم ومحتوى الاختبار	تناسب حاجات الطلاب في صف مدرسي معين، ومرونة النواتج والمحتوى تسمح بتكييف الاختبار وفقاً لأهمية كل وحدة من الوحدات التعليمية	تعتمد على أهداف عريضة لفرقة دراسية معينة تشترك فيها معظم المدارس. ويتم معاينة المحتوى بطريقة منتظمة

نوعية المفردات	يقوم معلم واحد ببناء المفردات وعادة تكون نوعيتها محدودة لقلة خبرة المعلم بفنيات كتابة الأنواع المختلفة من المفردات، وضيق وقته.	تُبنى المفردات بواسطة مجموعة من الخبراء، وتخضع لعمليات تدقيق ومراجعة وتحرير، وتجريب ميداني، وتحليل إحصائي لكي تكون المفردات ذات نوعية عالية الجودة.
طرق التطبيق والتصحيح الثبات	يحددها المعلم، وتكون موحدة في الصف المدرسي الواحد، ولكن يمكن تعديلها بما يتفق وحاجات الطلاب. ثبات الدرجات لا يكون عادة معلوماً، ويمكن أن يكون مرتفعاً إذا اعتنى المعلم ببناء الاختبار وفق خطة مدروسة	يحددها ناشري الاختبارات، ويجب اتباع تعليمات الاختبار في هذا الشأن بمنتهى الدقة. ثبات الدرجات يكون مرتفعاً، ويتراوح عادة بين (0.8، 0.95)، وربما يزيد عن ذلك .
المعايير	لا تُبنى معايير Norms لهذه الاختبارات، ولكن يمكن للمعلم حساب المعايير للصف الذى يقوم بالتدريس فيه.	تُبنى معايير لهذه الاختبارات لمجموعات مختلفة من الطلاب، وتُفسر الدرجات الخام في ضوئها، ويستخدمها جميع المعلمين لمقارنة أداء طلاب صف معين بمختلف المجموعات العمرية والصفوف المدرسية على المستوى القومي أو المحلي.
الأغراض والاستخدامات	تناسب قياس أهداف تعليمية يحددها المعلم، ويقارن في ضوئها أداء الطالب بأداء أقرانه في الصف المدرسي، أو تحديد ما اكتسبه الطالب من معارف ومهارات تتعلق بمحتوى دراسي معين.	تناسب بدرجة أفضل قياس الأهداف العريضة لمنهج دراسي معين، وتُستخدم هذه الاختبارات في المقارنة بين الصفوف والمدارس، واتخاذ قرارات تتعلق بالطلاب والمعلمين والمناهج المدرسية.

الفصل السادس

اختبارات الذكاء العام

- مقدمة
- مفهوم الذكاء وتطوراته التاريخية
- التعريفات المتعددة للذكاء
- نظريات بنية الذكاء
- تصنيف اختبارات الذكاء العام
- اختبارات الذكاء العام الفردية
- اختبارات الذكاء العام الجماعية
- تعقيب على اختبارات الذكاء العام

مقدمة:

تناولنا في الفصل السابق الاختبارات التحصيلية التي تُستخدم في قياس نواتج التعلم المتعلقة بالمناهج والمواد الدراسية، ومقارنة أداء الطالب بأداء أقرانه. غير أنه من الأهمية أيضاً مقارنة أداء الطالب بقدرته على التعلم لكى نتحقق من توظيفه لإمكاناته العقلية إلى أقصى حد ممكن. فالمعلمون يدركون أن جميع الطلاب يمكنهم إحراز تقدم دراسى والوصول إلى مستوى واحد، ولكنهم يعملون على توفير فرصة كافية لكل منهم لتحقيق أعلى مستوى مناسب لإمكاناته. وهذا يتطلب معرفة مستوى الذكاء العام للطالب واستعداداته الخاصة، وكيفية استخدام الاختبارات المناسبة لقياس إمكاناته العقلية استخداماً مستنيراً. ولتحقيق ذلك ينبغي على المعلم معرفة المقصود بمفهوم الذكاء والأطر النظرية والمنهجية التي تستند إليها اختبارات الذكاء العام، وأنواع هذه الاختبارات، ومحتواها، وكيفية تفسير درجاتها، واستخدام نتائجها استخداماً مناسباً.

لذلك يتناول هذا الفصل التعريفات المتعددة لمفهوم الذكاء في ضوء التطور التاريخى لحركة القياس العقلى، ونظريات بنية الذكاء، وتصنيف اختبارات الذكاء العام، وأمثلة لأهم اختبارات الذكاء العام الفردية والجماعية، وينتهى الفصل بتعقيب عام على هذه الاختبارات. وسوف أفرد الفصل التالي لاختبارات الاستعدادات الخاصة .

مفهوم الذكاء وتطوراته التاريخية :

يُعد مفهوم الذكاء من المفاهيم السيكولوجية التي نالت كثيراً من الخلاف والجدل بين علماء النفس وخبراء القياس منذ بداية القرن الماضي. ولعل معرفة التطور التاريخي لحركة القياس العقلى تلقى الضوء على هذا المفهوم باعتباره من المفاهيم التي تتميز بالتغير المستمر. فقد نشأ الاهتمام بضعاف العقول، والكشف عن الدرجات المختلفة للضعف العقلى في القرن التاسع عشر، وذلك بهدف وضع أسس لتصنيف الأفراد غير الأسوياء، والاهتمام بالتمييز بين مرضى العقول، وضعاف العقول. ويعد اسكرول Esquirol، وهو طبيب فرنسى، أول من حاول

تصنيف ضعاف العقول عام 1938 في مستويات مختلفة تبدأ من الأسوياء إلى أدنى مستويات الضعف العقلي. ولكن المعايير التي اعتمد عليها في هذا التصنيف لم تكن موضوعية، حيث اعتمد على شكل جمجمة الإنسان، وحجمها، ونسبتها إلى الجسم، وهذه تعد من المظاهر الجسمية، وبخاصة في تشخيص المستويات الدنيا للضعف العقلي. غير إنه اعتمد على بعض طرق قياس السلوك اللفظي في تشخيص المستويات العليا للضعف العقلي. وحاول بعض العلماء في ذلك الوقت وضع بعض الاختبارات التي تقيس المهارات الحسية الحركية والعملية لتدريب ضعاف العقول. وبالطبع لم تجدى هذه الطرق غير الموضوعية في الكشف عن ضعاف العقول أو تدريبهم (Hunt, 1961) .

وقد شهد منتصف القرن التاسع عشر تحولاً ملحوظاً في تطور حركة القياس العقلي، حيث تحوّل الاهتمام من التأملات الشخصية إلى التجارب.

فقد أسّس ڤنت Wundt عام 1879 أول مختبر لعلم النفس التجريبي في ليبتزج بألمانيا. وانصب اهتمام ڤندت وتلاميذه على بحث الظواهر النفسجسمية، مثل: الظواهر الحسية، وزمن رد الفعل. وفى أواخر القرن التاسع عشر حدث تقدم كبير في علم البيولوجي نتيجة لإسهام شارلز دارون بنظريته في التطور والنشوء والارتقاء.

ولعل العالم البيولوجي الإنجليزي فرانسيس جالتون Galton، وهو من أتباع دارون البارزين هو الذى وضع المعالم الرئيسة للقياس العقلي. فقد اهتم جالتون بمشكلة الوراثة، وأنشأ معمل دراسة الوراثة في لندن. وأجرى دراسات وثائقية من سجلات طلاب بعض المدارس الإنجليزية تتعلق بالمظاهر الجسمية والحسية لهؤلاء الطلاب، واستخلص أن الطلاب الذين يتميزون بقدرة عالية في التمييز الحسي يُعدون أذكى الطلاب بعامة.

وقد أدت جهود جالتون والأساليب الإحصائية التي استخدمها في ذلك الوقت إلى حفز تلاميذه من أمثال بيرسون Pearsoon، وكاتل Cattell، إلى الاهتمام بعلم الإحصاء كوسيلة للبحث النفسي. وقد عاد كاتل إلى الولايات المتحدة الأمريكية، حيث أشرف على معامل علم النفس التجريبي ونشر آراءه عن القياس العقلي. ولكنه اعترض على وجهة نظر جالتون، وهي أن قياس الوظائف العقلية يمكن الاستدلال عليه عن طريق قياس الوظائف النفسية البسيطة، مثل: التمييز الحسي، وقياس زمن رد الفعل، وغير ذلك من الوظائف الدنيا. لذلك تحوّل اهتمامه إلى

دراسة الفروق الفردية في نواحي أخرى مثل الذكاء. وقد قام ببناء مجموعة من الاختبارات العقلية التي اشتملت على مقاييس لسرعة حركة الذراع، أو التمييز بين الأوزان، وغير ذلك. إلا أنه تبين بعد ذلك أنه لا توجد علاقة بين درجات هذه الاختبارات ودرجات التحصيل الدراسي، مما أدى إلى دحض افتراض وجود ارتباط بين القدرة الحسية والذكاء. ومع هذا فإن هذه الفكرة أسهمت في اقتراح مقاييس جديدة للذكاء.

وقد انتقد بينيه **Binet**، وهو عالم نفس فرنسي، الاختبارات المستخدمة في ذلك الوقت لقياس الذكاء، واقترح هو وزميله سيمون **Simon** بناء اختبار جديد لقياس القدرة العقلية Men-tal Ability على أساس الوظائف العقلية العليا. ورأى بينيه أن الذكاء قدرة عقلية عامة تتكون من مجموعة من القدرات المنفصلة وهي: الاستدلال، والحكم الجيد، والفهم الجيد، واتخاذ توجه مناسب للتفكير والحفاظ على هذا التوجه، وتعديل التفكير بما يحقق هدفاً معيناً، والنقد الذاتي.

وقد استغرق بناء الاختبار حوالي عشر سنوات، وظهر الاختبار أو المقياس الجديد عام 1905، حيث نال اهتماماً كبيراً من علماء النفس في العالم.

وقام تيرمان Terman وهو عالم نفس أمريكي، بمراجعة الاختبار وترجمته إلى اللغة الإنجليزية في جامعة ستانفورد بالولايات المتحدة الأمريكية، وأطلق عليه اسم: اختبار ستانفورد – بينيه للذكاء.

ويرجع الفضل في تقرير أول وحدة في القياس العقلي، وهي "العمر العقلي Mental Age (MA)" إلى بينيه.

بينما يرجع الفضل إلى تيرمان في إدخال مصطلح جديد، وهو مصطلح "نسبة الذكاء Intelligence Quotient (IQ)". وعلى الرغم من انتشار استخدام اختبار بينيه في قياس الذكاء العام، إلا أنه تبين فيما بعد أنه مشوب بأوجه قصور متعددة. لذلك قام دافيد ويكسلر Wechsler عام 1939 ببناء اختبار ذكاء فردي يختلف عن اختبار بينيه في أنه يقيس بدرجة أساسية ذكاء الراشدين، والتعرف على الأنماط المنحرفة للاستجابات التي تميز مختلف أنواع الأمراض النفسية. غير أن نتائج الدراسات النفسية لم تستطع التوصل إلى أنماط معينة من

الاستجابات تميز الأفراد الذين يعانون من مشكلات نفسية مختلفة. لذلك أصبح يقتصر استخدام اختبار ويكسلر على قياس الذكاء، وليس تشخيص الأمراض النفسية.

ومن التطورات الأخرى التي حدثت في حركة القياس العقلي هي ظهور الاختبارات الجماعية Group Tests لقياس الذكاء. فمن المعلوم أن اختبار بينيه اختبار فردي Individual Test، أي يطبق على فرد واحد في وقت معين، ويحتاج إلى شخص على درجة عالية من التدرب ليشرف على تطبيقه وتفسير نتائجه.

ولكن عندما دخلت الولايات المتحدة الامريكية الحرب العالمية الأولى عام 1917 تطلب الأمر تصنيف الأفراد وفقاً لذكائهم بغرض توجيههم نحو الوظائف العسكرية المناسبة لقدراتهم.

لذلك شرع أحد الخبراء النفسيين في الجيش الأمريكي وهو آرثر اوتيس Arther Otis مع مجموعة من زملائه في بناء مجموعة من الاختبارات الجماعية لقياس القدرة العقلية للمجندين في الجيش الأمريكي، أي تطبق على مجموعات كبيرة من الافراد في الوقت نفسه. ومن اشهر هذه الاختبارات: اختبار ألفا Alpha Test، وهو يقيس القدرة العقلية للأفراد غير الناطقين باللغة الإنجليزية أو الأميين.

وبعد انتهاء الحرب العالمية الأولى انتشرت هذه الاختبارات، وترتب على ذلك تطور سريع لحركة القياس العقلي في الولايات المتحدة الأمريكية، وفي كثير من دول العالم. ومن أهم هذه التطورات بناء اختبارات لقياس مختلف مظاهر السلوك، واختبارات جماعية لقياس الذكاء في مختلف الأعمار بما في ذلك الأطفال والراشدين، وبطاريات لقياس الاستعدادات الخاصة، وتبع ذلك بناء مقاييس للسمات غير المعرفية، مثل: استبيانات الشخصية، والاختبارات الإسقاطية، وغيرها من طرق قياس الشخصية التي سنتناولها في الفصل الثامن.

التعريفات المتعددة للذكاء:

تتعدد وجهات نظر علماء النفس فيما يتعلق بتعريف مفهوم الذكاء Intelligence، لذلك لا يوجد اتفاق على تعريف موحد يمكن الاستناد إليه في قياس الذكاء. ولعل التطور التاريخي لحركة القياس العقلي الذي أوجزناه يوضح ذلك. فمحاولات تعريف الذكاء تواجه مشكلة اعتبار أن الذكاء له كينونة حقيقية بحيث يمكن تحديدها وقياسها. فالذكاء هو مجرد مصطلح يدل على مجموعة من السلوك المهم في حياة الأفراد. لذلك لا ينبغي النظر إلى

مصطلح أو كلمة الذكاء على أنها بديلاً للشيء الذي تمثله، أي السلوك، ومن ثم تحديد الشيء بدلاً من تحديد السلوك. فمصطلح الذكاء مثل مصطلح "الطقس" هو كلمة تصف ظاهرة تسلك بطريقة معينة. فنحن لا نقيس الذكاء، ولا نقيس الطقس، وإنما نقيس السلوك في كل من الحالتين، وليس المفهوم أو المصطلح ذاته. غير أننا نستدل على المفهوم من نتائج قياس السلوك، ونحكم على الفرد بأن ذكاءه مرتفع أو منخفض، أو أن قدرته اللفظية أفضل من قدرته غير اللفظية.

لذلك تحاول كثير من التعريفات وصف الذكاء وكيف يعمل. ولعل هذا يتضح في تعريف بينيه للذكاء، حيث أشار إلى كيفية إحداثه لأثر في السلوك، وخصائصه الأساسية.

وكذلك اقترح ثورنديك **Thorndike** تعريفاً للذكاء بأنه قدرة الفرد على القيام بمهام معينة مقارنة بأقرانه. وأداء هذه المهام ونتاجها هو الذي يعكس الذكاء.

وقد اقترح بيرت **Burt** التمييز بين التعريف الاسمي Nominal Definition، والتعريف الفعلي Real Definition للذكاء؛ فالتعريف الاسمي يشير إلى كيفية استخدام مفهوم الذكاء، بينما يشير التعريف الفعلي إلى تفسير طبيعة الذكاء. ولهذا تبدو أهمية تعريف المفهوم تعريفاً اسمياً قبل محاولة قياس الذكاء. أما التعريف الفعلي للذكاء فإنه يتطلب مزيداً من البحث والدراسة.

غير أن المتتبع للتعريفات الحالية للذكاء يلاحظ تنوع السلوك الذي يشير إليه التعريف، مما ينعكس على المهام الاختبارية. فاختبار بينيه يتضمن مهاماً فرعية متعددة، ولكنها تضم جميعاً في درجة واحدة تُعبر عن القدرة العقلية العامة. بينما تشتمل اختبارات ويكسلر على موازين فرعية Subscales متعددة تؤدي إلى نسب ذكاء لفظية وأخرى غير لفظية، وكذلك نسب ذكاء كلية. أما جيلفورد **(Guilford, 1983)** فقد اقترح بنية تنظيمية للذكاء ثلاثية الأبعاد، هي: المحتوى، والعمليات، والنواتج. في حين أن هوارد جاردنر **(Gardner, 1983)** اقترح نظرية الذكاءات المتعددة Multiple Intelligences. وقد انتقد روبرت سترنبرج **(Sternberg, 1988)** نظرية جاردنر، واقترح نظرية أخرى تعتمد على النظرية الثلاثية Triarchic Theory، حيث تميز بين العمليات الداخلية، والخبرة، والعلاقة بينها وبين متطلبات العالم الخارجي. وقد أكدت هذه النظرية أهمية مفهوم الضرورة Necessity إلى جانب مفهوم الذكاء؛ فالقدرات التي يحتاجها الفرد من أجل الحياة تعد جزءاً من الذكاء، اما القدرات الأخرى فيمكن اعتبارها قدرات خاصة.

وقد بينت نتائج كثير من البحوث أنه على الرغم من تعدد القدرات أو المهارات العقلية الخاصة، الا انها مترابطة، مما يعزز بدرجة معينة ما سبق تسميته العامل العام "g" Factor، أو ما يُعرف بالقدرة العقلية العامة. غير أن بعض الافراد يتميزون في إحدى القدرات المتعلقة بمفهوم الذكاء. وربما يختلف ذلك عن وجهات النظر المبكرة التي كانت ترى أن الذكاء قدرة غير متغيرة. فهناك الآن أدوات تستند إلى فكرة تفاعل السمات والمعالجات، أو تفاعل القدرات والمعالجات، مثل: استبيانات الانماط المعرفية، واختبار السلوك التكيفي الذي يقيس قدرة الفرد على أداء مهام تعلم في الحياة اليومية.

والخلاصة أن هذه التعريفات تؤكد أن الذكاء ليس قدرة فطرية كامنة Capacity، وإنما سمة سلوكية تعتمد إلى حد ما على التعلم السابق والسلوك المكتسب. وينبغي أن يقيس اختبار الذكاء سلوكاً مكتسباً يتفق علماء النفس على تسميته ذكاءً، وعندئذ يصبح الاختبار بمثابة تعريف إجرائي للذكاء. وعلى الرغم من اختلاف اختبارات الذكاء في انماط السلوك الذي تقيسه، إلا أن الارتباط بين درجات هذه الاختبارات مرتفع نسبياً.

نظريات بينيه الذكاء:

اتضح من عرضنا للتطور التاريخي لحركة القياس العقلي أن بينيه Binet وسيمون Simon قاما بتطوير أول اختبار فردي للذكاء لقياس مستوى الذكاء العام لدى الفرد. وعلى الرغم من أن كثيراً من علماء النفس الآخرين كانت تصوراتهم الفكرية عن الذكاء انه خاصة أو قدرة عقلية عامة، إلا أن هناك نظريات أخرى معارضة لهذه التصورات.

ويعد سبيرمان (Spearman, 1927) من العلماء الذين اعتبروا الذكاء مكون من عاملين، ولذلك أطلق على نظريته في الذكاء "نظرية العاملين Two Factor Theoy"؛ فقد اقترح أن الذكاء يشتمل على عامل عام (General Factor (g، وعوامل نوعية Specific Factors .

اما ثيرستون (Thurstone 1944) فقد قام بتطوير نظرية العوامل المتعددة Multiple Factors استند إليها في بناء اختبار القدرات العقلية الأولية الذي يقيس سبعة عوامل، هي: القدرة على الطلاقة اللفظية، القدرة على الفهم اللغوي، القدرة العددية، القدرة المكانية، السرعة الادراكية، القدرة على التذكر، القدرة على الاستدلال الاستقرائي.

أما ڤيرنون (Vernon, 1961) فقد اقترح بنية هرمية للقدرات العقلية تبدأ بالعامل العام الذي يتفرع إلى عاملين طائفيين رئيسين، هما: عامل لفظي – تربوي، وعامل حركي – ميكانيكي.

وهذين العاملين الطائفيين أقل عمومية من العامل العام (g) الذي اقترحه سبيرمان، ولكنهما اكثر عمومية من العوامل التي اقترحها ثيرستون.

واقترح كاتل (Cattell, 1963) نظرية توضح أن الذكاء يشتمل على مكونتين إحداهما الذكاء السائل Fluid Intelligence، والأخرى الذكاء المتبلور Crystallized Intelligence. ويعد الذكاء السائل اكثر عمومية، وارتفاع مقداره يدل على أن الفرد يستطيع القيام بمهام كثيرة ومتنوعة. لذلك فإن الذكاء السائل هو كفاءة عقلية تتميز بالتجريد وغير لفظية بدرجة أساسية، ومتحررة من الثقافة. أما الذكاء المتبلور فهو اكثر صلة بالثقافة أو البيئة، ويمثل قدرة الفرد على إنجاز مهام أكثر نوعية وارتباطاً بالثقافة، وبذلك فإن هذه القدرة تماثل التحصيل.

واقترح جيلفورد (Guilford 1959, 1969) نموذجاً لبنية العقل (Structure of Intellect (SI، حيث اعتبر أن الذكاء يتكون من عوامل متعددة قام بتصنيفها في أبعاد ثلاثة رئيسة هي: العملية التي يجريها الفرد، ونوع المادة أو المحتوى، ونوع النتاج. وكل بعد من هذه الأبعاد أعيد تقسيمه إلى خمس عمليات، وأربعة أنواع من المحتوى، وستة أنواع من النواتج. ويمكن تمثيل هذه الأبعاد بمكعب يحتوي على 4 x 5 x 6 = 120 خلية، يمثل كل منها مظهراً معيناً من مظاهر النشاط العقلي. واستطاع جيلفورد أن يبين وجود (82) عاملاً من هذه العوامل. واقترح قياس كل منها على حدة. وقد نال هذا النموذج اهتماماً كبيراً من جانب علماء النفس باعتباره نموذجاً نظرياً منطقياً أكثر من كونه نموذجاً عملياً تطبيقياً، حيث إنه ربما يصعب بناء اختبارات تناظر كل خلية من خلايا المكعب، وحتى إذا أمكن ذلك فإن هذه الاختبارات لا يكون لها فائدة حقيقية من حيث قيمتها التنبؤية، نظراً لأنها تقيس عوامل نوعية بدرجة كبيرة. ومن الجدير بالذكر أن جيلفورد اعتمد في هذا النموذج على أسلوب التحليل العاملي Factor Analysis، وهو من الأساليب الإحصائية متعددة المتغيرات الذي يهدف لاختزال عدد كبير من المتغيرات وتجميعها في عدد قليل من العوامل التي تفسر التباين المتسق في هذه المجموعة من المتغيرات. وهذه العوامل Factors تعد تكوينات فرضية Constructs تنطوي تحتها مجموعة المتغيرات وتفسرها.

أما جاردنر (Gardner, 1989) وهو من علماء النفس المعرفي بجامعة هارفرد فقد حاول خلال العقدين الماضيين تطوير نظرية الذكاءات المتعددة Multiple Intelligences وكيفية الاستفادة منها في التعليم. وأوضح عدداً من هذه القدرات والمهارات المتمايزة التي يمتلك كل فرد درجات

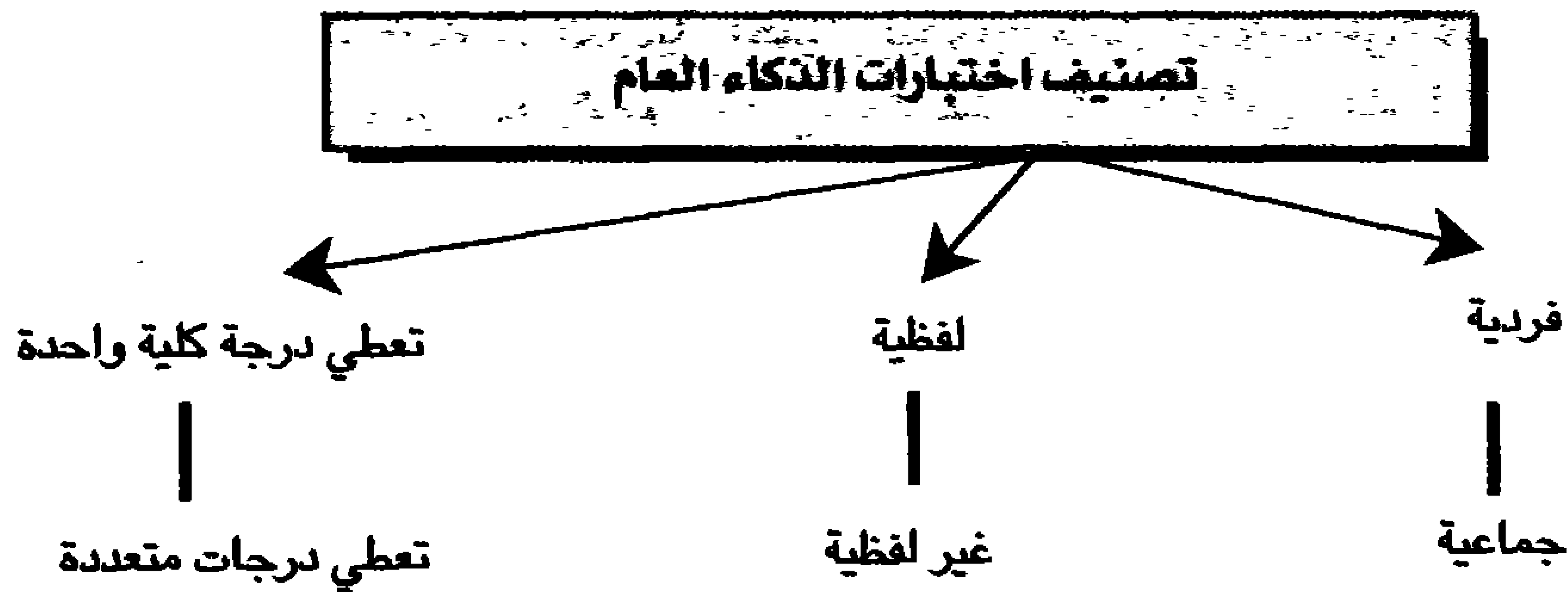
معينة من كل منها، وهي: القدرة البصرية – المكانية، القدرة الجسمية – الحركية، القدرة الموسيقية – الإيقاعية، القدرة على التفاعل بين الأفراد، القدرة على معرفة مشاعر الفرد الذاتية، القدرة المنطقية – الرياضية، القدرة اللفظية – اللغوية.

وأوضح جاردنر أن هذا يخالف الاعتقاد بأن الذكاء الانساني سمة ثابتة واحدة، وانما توجد ذكاءات متعددة يمكن تنميتها بأساليب تعليم مناسبة، وتتشكل بتفاعل الفرد مع بيئته تفاعلاً ايجابياً مثمراً.

ومن هذا يتضح عدم اتفاق علماء النفس حول بنية الذكاء، لذلك تعددت الاختبارات التي يُطلق عليها اختبارات الذكاء، وتنوعت تنوعاً كبيراً، وربما لا تقيس جميعها الشيء نفسه. لذلك ينبغي عند انتقاء أحد هذه الاختبارات وتفسير درجاتها مراعاة تعريف الذكاء الذي استند إليه بناء الاختبار. ولمزيد من تفاصيل نظريات بنية الذكاء يمكن الرجوع إلى كتاب "القياس والتقويم التربوي والنفسي" للمؤلف.

تصنيف اختبارات الذكاء العام:

سبق أن أوضحنا في الفصل الأول أنواع الاختبارات التربوية والنفسية وفقاً لتصنيف يستند إلى عدد من الأبعاد. وينطبق هذا التصنيف في كثير من أبعاده على اختبارات الذكاء، حيث إن هذه الاختبارات يمكن أيضاً تصنيفها بطرق متعددة. ولعل التصنيف الذي يوضحه الشكل التخطيطي (6 - 1) التالي يفيد في التعرف على أنواع اختبارات الذكاء المتوافرة، وانتقاء الاختبار المناسب لغرض معين:



شكل (6 - 1): يوضح تصنيف اختبارات الذكاء العام

الاختبارات الفردية والجماعية:

تنقسم اختبارات الذكاء إلى إختبارات فردية، واختبارات جماعية. والاختبار الفردي Individual Test يقوم بتطبيقه اخصائي نفسي على فرد واحد، ويقوم الأخصائي بتقديم المهام التي يشتمل عليها الاختبار للفرد المختبر شفوياً وجهاً لوجه، ويستجيب الفرد بالإشارة إلى مثير معين، أو تقديم إجابة شفوية، أو التعامل يدوياً مع مهمة معينة، ويتطلب تطبيق الاختبار اخصائي نفسي متمرس على هذا النوع من الاختبارات الفردية، حيث أن الإخصائي يعد جزءاً من الإجراءات المقننة للاختبار. إذ ينبغي أن يقنن تعامله مع المختبر وفقاً لتعليمات الاختبار المحددة. وكثير من مفردات الاختبارات الفردية لا يتم تقدير درجاتها على أنها صحيحة أو خطأ، وانما ربما تكون لها إجابات متعددة مقبولة. لذلك تقدر عادة درجات مختلفة بحسب درجة صحة الإجابة. ويحدد في دليل الاختبار الإجابات التي يمكن اعتبارها صحيحة، والدرجة الجزئية التي تقدر لكل منها. وينبغي أن يلتزم الأخصائي بهذه التعليمات ولا يحيد عنها، لكي لا تتأثر الدرجات بأحكامه الذاتية عن ذكاء المختبر.

لذلك يستغرق تطبيق وتصحيح هذا النوع من الاختبارات وقتاً طويلاً نسبياً، ويصلح بدرجة افضل للأطفال الصغار، وكذلك في المواقف الكلينيكية، حيث تسمح للأخصائي بالملاحظة الدقيقة للمختبر. وتتميز هذه الاختبارات بأن تقديراتها للذكاء تكون أكثر صدقاً لأنها تطبق في ظروف مثلى. فالمختبر تكون دافعيته للإجابة وتعاونه أفضل، ويتأكد الأخصائي بأن المختبر في صحة جيدة، وغير متوتر، وانتباهه مركزاً على المهام المعطاة، ومستوعباً استيعاباً كاملاً للتعليمات. كما أن قدرته على القراءة لا يكون لها تأثير على أدائه أو درجته في الاختبار.

ومن أهم أمثلة الاختبارات الفردية: اختبار بينيه، واختبارات ويكسلر للذكاء التي سأتناولها بالتوضيح بعد قليل.

أما الاختبارات الجماعية Group Tests فقد برزت كما سبق أن أوضحنا نتيجة لحاجة ملحة، وهي قياس القدرة العقلية للجنود في الحرب العالمية الأولى، حيث يصعب بدرجة كبيرة تطبيق اختبار فردي مثل اختبار بينيه على أعداد كبيرة من هؤلاء الجنود.

وتشتمل الاختبارات الجماعية عادة على مفردات تتطلب إجابتها الورقة والقلم. لذلك يمكن تطبيقها على مجموعة كبيرة من الأفراد في وقت واحد. مما يوفر الوقت والجهد والكلفة. ولا تتطلب هذه الاختبارات من القائم بتطبيقها تدريباً متخصصاً، حيث إنه يقرأ تعليمات الاختبار،

ويسمح ببعض الوقت لكي يألف المختبرون كيفية الإجابة على مفردات الاختبار، ويعطي لهم الإذن ببدء الإجابة، ويجمع أوراق الإجابة عند انتهاء الزمن المخصص للاختبار.

ويتميز هذا النوع من الاختبارات بسهولة تطبيقه، والانخفاض النسبي لكلفة تصحيح الاختبار، وكذلك سرعة التصحيح. غير أنه في مقابل ذلك تكون ظروف تطبيق الاختبار غير مقننة بدرجة كاملة، وربما تتأثر الدرجات بعوامل مختلفة، مثل: سرعة القراءة، والخلفية التعليمية للمختبرين. لذلك فإن درجات اختبارات الذكاء الجماعية تقلل عادة من تقديرات ذكاء المختبرين.

ومع هذا، فإن هذه الاختبارات تناسب الأفراد الراشدين العاديين في المواقف الجماعية مثل: الصفوف المدرسية، والمؤسسات الصناعية والعسكرية.

ومن أفضل أمثلة الاختبارات الجماعية: اختبار أوتيس للقدرة العقلية (اختبار ألفا) الذي أشرنا إليه من قبل، واختبار ثيرستون لليقظة العقلية الذي يزودنا بدرجتين فرعيتين للذكاء إحداهما درجة للمفردات الكمية، والأخرى للمفردات اللغوية، وكذلك درجة كلية. وهناك اختبار التصنيف العام (AGCT) الذي تم تقنينه على مجتمع من المجندين بالجيش الامريكي أثناء الحرب العالمية الثانية، وأصبح يُستخدم في قياس ذكاء الأفراد المدنيين. ويركز على الاستدلال العددي، والاستدلال اللفظي، والادراك المكاني (القدرة على تصور الأشكال في الفراغ).

وزمن تطبيق هذا الاختبار 40 دقيقة ويؤدي إلى درجة كلية محولة، ويعد من اختبارات الذكاء العام الجماعية التي تتميز بالصدق. ويُستخدم في انتقاء الأفراد، ولكنه أقل فائدة في التنبؤ بنجاح الطلبة في الكليات، وذلك لأن نوع الذكاء الملائم للنجاح في الكليات يتطلب قدرات خاصة. لذلك فإن الاختبار الشائع الاستخدام في هذا المجال هو اختبار القبول بالجامعات (CEEB) الذي أعده مركز الخدمات الاختبارية بالولايات المتحدة الامريكية (ETS)، يشتمل على نوعين من الاختبارات أحدهما اختبار الاستعداد العام، والثاني اختبارات تحصيلية في مختلف المواد الدراسية. ويفيد هذا الاختبار في التنبؤ بنجاح الطلاب في المرحلة الجامعية الأولى.

الاختبارات اللفظية وغير اللفظية:

تستخدم معظم اختبارات الذكاء الجماعية، وبخاصة التي تناسب الراشدين أنواعاً معينة

من المفردات مثل: مفردات لغوية، إكمال جمل، متشابها لفظية، حساب، متضادات، تسلسل لفظي، تمييز الفروق، تفسير أمثال، ترتيب جمل، استدلال، اتباع تعليمات، استنتاج مترتبات، علاقات بين اشكال، متشابهات شكلية، متسلسلات عددية، تعامل مع مساحات، وغير ذلك. لذلك تعتمد هذه الاختبارات على القراءة والكتابة والتعبير الشفوي، ويطلق عليها الاختبارات اللفظية Verbal Tests. أما اختبارات الذكاء غير اللفظية Nonverbal Tests فإن اللغة لا تلعب دوراً في الإجابة؛ فهذه الاختبارات تستخدم الصور والأشكال أو اي مواد رمزية، لذلك فإنها تناسب بدرجة أفضل الأطفال الصغار والأميين لقياس الجوانب غير اللفظية للذكاء، مثل الادراك المكاني. أما الاختبارات غير اللغوية Non language فهي لا تعتمد على أي شكل من أشكال التواصل الكتابي أو الشفوي أو القرائي. وإنما يقدم الفاحص التعليمات من خلال الإيماءات والإشارات، ويستجيب الفرد بالإشارة إلى الأشياء أو التعامل معها يدوياً مثل: الكتل الخشبية، والمتاهات، والصور، وغير ذلك. وتطبق هذه الاختبارات على الأطفال الصم، أو الذين لا يمكنهم التواصل بأي شكل من أشكال اللغة العادية.

غير أن بعض اختبارات الذكاء تشتمل على أجزاء لفظية وأجزاء غير لفظية ميزات هذا النوع من الاختبارات انه يمكن باستخدامه تحديد قدرة الفرد على التعلم في مستويين مختلفين. فدرجته في الاختبار اللفظي تفيد في التنبؤ بنجاحه الدراسي، وذلك لأن القدرة اللفظية تلعب دوراً مهماً في تعلم المهام الدراسية، أما درجته في الاختبار غير اللفظي فتقدم تقديراً افضل لقدرته الكامنة على التعلم التي تبدو منخفضة. فمثلاً الطالب الذي قدرته العقلية تفوق المتوسط ولكن قدرته على القراءة منخفضة يُتوقع أن يحصل على درجة منخفضة في الجزء اللفظي من الاختبار، بينما ربما يحصل على درجة مرتفعة في الجزء غير اللفظي. وعلى الرغم من أن احتمال نجاحه الحالي في الدراسة يكون منخفضاً نظراً لانخفاض قدرته على القراءة، إلا أنه يُحتمل نجاحه المستقبلي اذا قدم له تعليماً علاجياً مناسباً.

غير أن الفروق بين درجات الجزء اللفظي، والجزء غير اللفظي ينبغي أن تكون كبيرة نسبياً قبل استخدامها في اغراض التشخيص، وذلك لارتباط كل منها بالآخر ارتباطاً مرتفعاً إلى حد ما، لاشتراكهما في عامل القدرة العقلية العامة.

الاختبارات التي تعطي درجة كلية واحدة، أو تعطى درجات متعددة:

بعض اختبارات الذكاء التي ينتج عنها درجة واحدة كلية تُصمم بحيث تقيس القدرة العقلية

العامة. لذلك فإن الجوانب النوعية لهذه القدرة، مثل: الجوانب اللفظية، والعددية، والاستدلال المجرد تُضم معاً في قياس واحد للقدرة العقلية. ويتم ترتيب مفردات هذا النوع من الاختبارات التي ينتج عنها درجة واحدة كلية في نمط حلزوني Spiral - Omnibus Pattern، أي خلط الأنواع المختلفة من المفردات، وترتيبها ترتيباً تصاعدياً من حيث صعوبتها. وبذلك لا يكون هناك اختبارات فرعية، مما يجعل تعليمات الاختبار مجموعة واحدة، ويكون زمن الاختبار قصيراً نسبياً. ومن أمثلة هذا النوع من اختبارات الذكاء: اختبارات القدرة العقلية لهلمون، ونيلسون Helmon - Nelson Tests of Mental Ability للصفوف من الحضانة إلى نهاية المرحلة الثانوية، واختبار القدرة المدرسية لأوتيس ولينون Otis - Lennon School Ability Test للصفوف من الأول الابتدائي إلى نهاية المرحلة الثانوية.

وتستخدم عادة الاختبارات التي تعطى درجة واحدة كلية في قياس الاستعداد للتعلم المدرسي، والتنبؤ بالنجاح في التعليم المدرسي في المستقبل. ومفردات هذه الاختبارات تكون لفظية مما يزيد من قيمتها التنبؤية، وذلك لأن طبيعة التعلم المدرسي يكون لفظياً.

أما اختبارات الذكاء التي تعطى درجات متعددة فإنها تقيس عدداً من القدرات العقلية، ومثال ذلك اختبار القدرات المعرفية Cognitive Abilities Tests التي تناسب التلاميذ من الحضانة إلى نهاية المرحلة الثانوية. وتُستخدم هذه الاختبارات في التنبؤ التمايزي. وفي بطاريات الاختبارات التي ينتج عنها درجات متعددة يكون هناك تداخل بين الاختبارات الفرعية يعكس القدرة العقلية العامة، وكذلك القدرات النوعية التي تقيسها هذه الاختبارات.

اختبارات الذكاء العام الفردية

مقياس ستانفورد - بينيه للذكاء العام:

يعد هذا المقياس أكثر مقاييس الذكاء العام انتشاراً واستخداماً، وقد مر هذا المقياس بمراجعات وتعديلات متتالية منذ الطبعة الأولى لمقياس بينيه – سيمون الأصلي عام 1916. وقد اجريت هذه المراجعات في الولايات المتحدة الأمريكية، ولعل أحدثها طبعة عام 1984. ويُستخدم هذا المقياس في قياس الذكاء العام للأعمار من عامين إلى الرشد، على الرغم من أنه يُستخدم بدرجة أكبر مع الأطفال من عمر عامين إلى ثمانية اعوام، ويعتمد المقياس على معايير عمرية للأداء. أي ترتب المجموعات المنفصلة من المفردات بحيث يكون مستوى صعوبتها

مناسباً للمستوى العمري المصممة له. فمجموعة المهام التي تقدم للأعمار من (2) إلى (5) أعوام يفصل بينها ستة أشهر عمرية، أما بالنسبة للأعمار من (5) إلى (14) عاماً فإنه يفصل كل منها عن الآخر عام واحد.

ويوجد اختبار للراشد المتوسط وثلاثة اختبارات للراشد المتفوق. ويستغرق تطبيق القياس عادة ما بين 50 , 75 دقيقة.

ويقيس هذا المقياس عدة مهارات يمكن تصنيفها في سبعة أقسام من المحتويات كالتالي:

- اللغة: تسمية الأشياء في صور، تعريف كلمات، تسمية كلمات مقفاة.
- الاستدلال: رسم توجهات، معرفة سبب سخافة عبارات لفظية.
- الذاكرة: تذكر جمل، تذكر أرقام.
- التفكير: شرح أمثال سائرة، توضيح أساس تماثل أشياء.
- الذكاء الاجتماعي: فهم علاقات اجتماعية، البحث عن سخافات في صور.
- الاستدلال العددي: إجراء تغييرات، حل مسائل رياضية حلولاً ذكية.
- النشاط البصري - الحركي: تكوين شكل معين على لوحة أشكال، رسم شكل معين عن طريق النقل. والمثيرات المستخدمة تشتمل على: كلمات، وأشياء، وصور. والاستجابات تتخذ أنماطاً متعددة، مثل: التحدث، الرسم، إجراء حسابات، الكتابة، وأنشطة حركية.

غير أنه يلاحظ أن هذه المهام السبعة تعتمد اعتماداً أساسياً على الاستدلال المجرد، ولا تعد القدرة على القراءة أمراً ضرورياً، ولكن يبدو أن المهارات اللفظية تلعب دوراً كبيراً في الإجابة على مفردات المقياس.

وهذا يضفي أهمية على مقياس ستانفورد - بينيه، لأن نسب الذكاء الناتجة عن المقياس تعد منبئاً جيداً بالنجاح في العمل المدرسي المعتاد.

ويتضمن المقياس مدى متسعاً من المستويات، حيث يشتمل على ستة اختبارات فرعية لكل من الأعمال المختلفة. ويبدأ الفاحص بتقديم المهام التي تقل مباشرة عن العمر الزمني للمتغير والتي يستطيع الإجابة عليها جميعاً إجابة صحيحة ثم ينقله إلى مهام أعلى إلى أن يصل إلى

المستوى الذي عنده يخفق في الإجابة على جميع الاختبارات الفرعية. وبذلك يمكن تحديد العمر العقلي القاعدي Basal Age، والعمر الأقصى Maximal Age .

وتتحدد درجة الإجابة الصحيحة لكل مهمة بمقارنتها بقائمة مطولة من إجابات متنوعة في دليل الاختبار. ومن ثم تحول الدرجة إلى ميزان عمري Age Scale وذلك بإضافة عدد معين من الشهور لكل إجابة صحيحة، وعندما تُجمع هذه الشهور على العمر القاعدي نحصل على العمر العقلي للمختبر.

ونظراً لصعوبة مساواة توزيع نسب الذكاء (IQ's) في مختلف المستويات العمرية، وعدم استقامة منحنيات النمو العقلي وافتقادها للمعنى بالنسبة لذكاء الراشدين، حيث إن هناك خلافاً على العمر الزمني الذي يمكن استخدامه في مقام نسبة الذكاء، فقد استُخدمت نسبة الذكاء الإنحرافية (DIQ) التي سبق إيضاحها في الفصل الرابع بدلاً من نسبة الذكاء المعتادة التي نحصل عليها بقسمة العمر العقلي على العمر الزمني، وضرب الناتج في (100)، وذلك في الطبعة المعدلة للمقياس عام 1960. ونسب الذكاء الانحرافية كما اوضحنا تعد درجات معيارية محولة متوسطها (100) وانحرافها المعياري (16) نقطة. ويمكن تفسير درجات الاختبار إما باستخدام نسبة الذكاء الإنحرافية (DIQ) التي تدل على المعدل النسبي للنمو العقلي للفرد مقارنة بأقرانه في المستوى العمري نفسه، أو باستخدام العمر العقلي الذي يدل على المستوى الحالي للنمو العقلي للفرد. وعلى الرغم من أن كلاً من نسبة الذكاء الانحرافية، والعمر العقلي تعد درجات محولة معيارية لانها تعتمد على أداء جماعة مرجعية أو عينة تقنين المقياس، الا أن العمر العقلي يمكن أيضاً تفسيره من حيث نوع وصعوبة المهمة التي يستطيع الفرد أداءها، اي يمكن تفسيره على أنه درجة متعلقة بمحتوى معين.

ويعد مقياس ستانفورد – بينيه من المقاييس التي تستند إلى أسس فنية، حيث اعتُني ببنائه وتقدير خصائصه السيكومترية. فمعظم الدراسات التي تناولته بالبحث توصلت إلى أن قيمة معاملات ثبات درجاته تزيد عن (0.90)، غير أن هذه القيمة تقل قليلاً بالنسبة لدرجات الأطفال الصغار. ويفيد هذا المقياس كغيره من مقاييس الذكاء في التنبؤ بالنجاح الدراسي المستقبلي، حيث إن معاملات الصدق التنبؤي والتلازمي باستخدام محكات، مثل: تقديرات الطلاب في المدارس، ودرجات الاختبارات التحصيلية تتراوح قيمها بين (0.40 ، 0.75) وذلك بحسب المستوى العمري للأفراد.

مقاييس ويكسلر للذكاء العام:

تعد مقاييس ويكسلر للذكاء Wechsler Intelligence Scales من المقاييس الفردية أيضاً التي تتميز بأنها حاولت تلافي عيوب مقياس ستانفورد – بينيه السابق. فمفردات مقياس بينيه Binet صُممت للأطفال، ولا تناسب الراشدين في معظم الأحيان، واستخدام مفهوم العمر العقلي والمعايير العمرية تبين محدودية فائدته للراشدين، كما أن معايير المقياس استندت إلى بيانات مستمدة من الأطفال وليس الراشدين. لذلك قام ويكسلر بتطوير مقياس خاص بذكاء الراشدين نُشر في صورته الأولى الذي أطلق عليها مقياس ويكسلر – بلفيو للذكاء – Wechsler Bellevue Scales عام 1939 نسبة إلى مستشفى بلفيو للأمراض النفسية بولاية نيويورك الامريكية الذي كان يعمل فيها خبيراً سيكولوجياً رئيساً. وكان الهدف من المقياس تجنب استخدام مفهوم العمر العقلي، وتقليل أثر عامل السرعة في الاستجابة الذي يعوق الراشدين، وتصميم مهام أكثر تشويقاً وملاءمة لهم، وكذلك زيادة الفائدة التشخيصية للمقياس باشتماله على مقاييس لفظية وادائية.

وقد أجريت مراجعة لهذا المقياس عام 1955 وأعيد تقنينه، وأصبح يسمى مقياس ويكسلر لذكاء الراشدين (WAIS) ويناسب الأفراد الذين يبلغون من العمر (16) عاماً أو أكبر. وتم ترتيب مفردات الاختبار في اختبارات فرعية منفصلة بدلاً من ترتيبها وفقاً للأعمار الزمنية. وأُجرى تحويل للدرجات الخام إلى نسب ذكاء انحرافية Deviation IQ (DIQ's)، ودُونت التحويلات في جداول المعايير.

ويستند المقياس إلى تعريف ويكسلر بأن الذكاء هو القدرة العامة للفرد للتصرف والسلوك الهادف، والتفكير المنطقي، والتعامل بفاعلية مع بيئته. وبذلك تبنّى ويكسلر مفهوم الذكاء العام، وان المقاييس المختلفة للذكاء ليست مقاييس لأنواع مختلفة من الذكاء، وأكد أن الذكاء ليس مجرد جمع قدرات تقيسها مفردات الاختبارات الفرعية للمقياس، وانما هو تكوين نسبي للقدرات بالإضافة إلى العوامل الدافعية والشخصية التي تؤدي إلى سلوك ذكي.

ويشتمل مقياس ويكسلر لذكاء الراشدين (WAIS) على (110) اختباراً فرعياً، تقيس ستة منها الجوانب اللفظية، والخمسة الباقية تقيس الأداء. وفيما يلي توضيحاً للاختبارات الفرعية لكل من القسمين:

■ المقياس اللفظي Verbal Scale :

(1) المعلومات (تذكر معارف): ويشتمل على أسئلة تتعلق بمعلومات عامة في موضوعات متعددة تناسب معظم الراشدين في المجتمع.

(2) الفهم (فهم المعلومات): ويقيس الحكم الواقعي والمعلومات، مثل: لماذا يدفع الناس الضرائب؟

(3) الحساب (القدرة العددية): مسائل كلامية تتطلب أقل قدر ممكن من العمليات الحسابية.

(4) التشابه (الاستدلال): ويشتمل على (13) زوجاً من الكلمات، ويُطلب من الفرد توضيح أوجه التشابه بين كل زوج منها.

(5) المفردات (القدرة اللفظية): تتطلب تعريف مفردات لغوية عامة.

(6) وسع الذاكرة للأرقام (الذاكرة): تتطلب إعادة ذكر أرقام بترتيبها الأصلي، أو بالترتيب العكسي.

■ مقياس الاداء Performance Scale :

(1) رموز الأرقام (الترميز): تتطلب مزاوجة الأعداد برموز مجردة.

(2) تكملة الصور (تحليل الاجزاء من الكل): تتطلب تسمية الجزء الناقص من صورة معطاة.

(3) تصميم المجسمات (تحليل مجسم معقد): يتطلب ترتيب أربعة مجسمات مختلفة الألوان لكي تكون تصميماً معيناً.

(4) ترتيب الصور (الاستدلال غير اللفظي): يتطلب ترتيب مجموعة من الصور بحيث تحكي قصة ذات معنى.

(5) تجميع الأشياء (توليف الكل من الأجزاء): يتطلب تجميع أجزاء متاهات تتباين في درجة تعقدها لكي تُكون شيئاً معيناً، وذلك باسرع ما يمكن.

ويستغرق تطبيق مقياس ويكسلر لذكاء الراشدين ساعة واحدة تقريباً. وتطبق الاختبارات الفرعية واحداً تلو الآخر، حيث تُقدم مفردات كل من هذه الاختبارات وفقاً لترتيب صعوبتها، ويتوقف التطبيق إذا اخفق الفرد في الإجابة عدداً محدداً من المرات. وبعض المفردات لها زمن محدد للإجابة، والبعض الآخر غير موقوت.

والقائم بتطبيق هذا المقياس ينبغي أن يكون على درجة عالية من التدرب والخبرة كما هو الحال في مقياس ستانفورد- بينيه، ويطبق المقياس بطريقة فردية. ويتم تقدير درجات المختبر بمقارنة إجاباته بقائمة تشتمل على بدائل الإجابات المقبولة، والمدونة في دليل المقياس، وتعتمد الدرجة على نوعية الإجابة. وتُحول الدرجات الخام التي يحصل عليها الفرد في كل اختبار فرعي إلى درجات معيارية متوسطها (10) وانحرافها المعياري (3) ومن ثم تجمع الدرجات المعيارية للاختبارات الفرعية، وتحول إلى ثلاث نسب ذكاء انحرافية، الأولى للمقياس اللفظي، والثانية لمقياس الأداء، والثالثة للمقياس ككل. وقد استخدم ويكسلر نسبة الذكاء الانحرافية (DIQ) التي متوسطها (100) وانحرافها المعياري (15) بدلاً من (16) المستخدمة في مقياس بينيه.

ونظراً لأن الأداء في المقياس يتدهور تدريجياً مع تقدم العمر، فإن نسبة الذكاء الانحرافية تم حسابها لكل مجموعة عمرية على حدة. وبذلك تدل هذه النسبة على مركز الفرد بالنسبة للجماعة المرجعية التي تشتمل على عينة من الأفراد في نفس المجموعة العمرية للفرد.

وقد أوضحت البيانات التي جمعها ويكسلر فيما يتعلق بالمقياس أن درجات المقياسين الفرعيين، والمقياس ككل تتميز بقدر مرتفع من الثبات يساوي أو يزيد عن (0.93)، كما أن درجات المقياس ككل ترتبط بدرجات مقياس ستانفورد – بينيه ارتباطا تبلغ قيمته حوالي (0.85). كما تتراوح قيمة الارتباط بين المقياس اللفظي ومقياس الأداء بين (0.77 ،0.81) في عينات ثلاث مختلفة من الأفراد (Wechsler, 1955) .

ومن الجدير بالذكر أن ويكسلر أعد مقياسين آخرين احدهما لذكاء الأطفال Wechsler Intelligence Scale for Children (Wisc) عام 1949 للأعمار من (5) اعوام إلى (15) عاماً. وأُجريت مراجعه على الاختبار، وأصبح يرمز له (WISC-R) ويناسب الأعمار من 6 اعوام ، 6 شهور إلى (16) عاماً. والآخر لذكاء أطفال ما قبل المدرسة.- Wechsier Pre school and Primary Scale of Intelligence (WPPSI) عام 1967 ويناسب الأطفال الذين تتراوح اعمارهم بين (4) اعوام، وستة أعوام ونصف.

واستند ويكسلر في بناء هذين المقياسين إلى الأسس التي استند إليها في بناء مقياس الذكاء للراشدين، حيث اشتمل كل من المقياسين المذكورين على مقياسين فرعيين أحدهما لفظي، ويحتوي على ستة اختبارات فرعية، والآخر أداء، ويحتوي على خمسة اختبارات فرعية.

اختبارات الذكاء الجماعية:

سبق أن أوضحنا أن الاختبارات الجماعية Group Test تتميز على الاختبارات الفردية في فعالية العملية الاختبارية وخفض كلفة تطبيق الاختبارات، حيث إنها تطبق على عدد كبير من الأفراد في وقت واحد. وتوجد أنواع متعددة من اختبارات الذكاء العام الجماعية، لعل أكثرها استخداماً الاختبارات التالية:

(1) اختبار أوتيس - لينون للقدرة العقلية (OLMAT): ويناسب المراحل من الحضانة إلى نهاية التعليم الثانوي، ويشتمل على ستة مستويات تناظر الصفوف الدراسية المختلفة. ويستغرق تطبيق اختبار المستوى الابتدائي حوالي (30) دقيقة، أما اختبارات المستويات الأخرى فيستغرق تطبيق كل منها حوالي ساعة واحدة. ويشتمل الاختبار في صورته الأخيرة على أنواع مختلفة من المفردات التي تُرتب حلزونياً أو دائرياً في الاختبار، وتتعلق بالجوانب اللفظية والتعليمية للعامل العام (g) في نظرية سبيرمان لبنية الذكاء. وينتج عن الاختبار درجة خام كلية واحدة يتم تحويلها إلى نسبة ذكاء انحرافية (DIQ)، ورتب مئينية للصفوف المدرسية، وتساعيات معيارية. واقتصرت عينة تقنين الاختبار على مجتمع طلاب المدارس فقط. وتتراوح قيم معامل ثبات درجات الاختبار بطريقة الصيغ المتكافئة بين (0.83) إلى ما يزيد عن (0.90)، وتزداد قيمة معامل الثبات بزيادة العمر الزمني للأفراد، وهناك بيانات للتحقق من صدق محتوى، وصدق التكوين الفرضي للاختبار.

(2) اختبار لورج - ثورنديك للذكاء (LTIT) : وهو من الاختبارات متعددة المستويات أيضاً من الحضانة إلى نهاية المرحلة الثانوية. ويستغرق تطبيقه حوالى ساعتين. وينتج عنه درجات للجانبين اللفظى، وغير اللفظى، وكذلك درجات كلية. ويتميز هذا الاختبار بأن ثبات درجات الجانب اللفظى يتراوح بين(0.83 ،0.91)، وثبات درجات الجانب غير اللفظى يتراوح بين (080 ، 0.88) باستخدام طريقة الصيغ المتكافئة. ويتم تحويل الدرجات الخام إلى نسبة ذكاء انحرافية لكل مستوى على حدة. ويتراوح مقدار الارتباط بين درجات هذه الاختبارات، ودرجات الاختبارات التحصيلية بين(0.60، 0.70).

(3) اختبار القدرات المعرفية (CAT) : ويعد أيضاً من الاختبارات متعددة المستويات من (3)

أعوام إلى نهاية المرحلة الثانوية، وتوجد طبعة منه تناسب الأعمار الأقل من (3) أعوام. وتشتمل على ثلاث بطاريات منفصلة يستغرق تطبيق كل فيها (35) دقيقة بالإضافة إلى زمن التعليمات الذى يبلغ حوالى (15) دقيقة. وتشتمل هذه البطاريات على جوانب لفظية (مفردات لغوية، إكمال جمل، تصنيف لفظي، متشابهات لفظية)، وجوانب كمية (سلاسل أعداد، علاقات كمية، تكوين معادلات)، وجوانب غير لفظية (تصنيف أشكال، متشابهات شكلية، تكوين أشكال) .

أما الطبعة الخاصة بالأعمار الأقل من (3) أعوام فإنها تستخدم مواداً بصرية، وتعليمات شفوية، وتتضمن اختبارات فرعية تشتمل على مفردات لغوية شفوية، ومفاهيم علاقية، ومفاهيم كمية. ويطبق هذا الاختبار مفرده تلو الأخرى لضمان عدم تأثره بسرعة الإجابة.

وقد أُعدت معايير مئينية، وتساعيات معيارية، وغيرها من المعايير المستمدة من عينه قوامها 200, 000 طالب من مختلف الصفوف المدرسية، ويتخطى معامل ثبات درجاتها القيمة (0.90).

غير أن الارتباط بين درجات اختباراته الفرعية تتراوح بين (0.40، 0.77)، ولم يوضح دليل الاختبار بيانات كافية للتحقق من صدقه.

(4) اختبار القدرات العقلية الأولية (PMA): أعد ثيرستون هذا الاختبار استناداً إلى نظريته التي سبق ذكرها في مستهل هذا الفصل والتى تُعرف بنظرية العوامل المتعددة للذكاء. ويعد هذا الاختبار بمثابة عدة بطاريات تناسب المراحل: الحضانة والصف الأول الإبتدائى، والصفوف من الثانى إلى الرابع، ومن الرابع إلى السادس، ومن السادس إلى التاسع، ومن التاسع إلى نهاية المرحلة الثانوية. وتقيس خمس قدرات هى :

معانى الكلمات، التعامل مع الأعداد بيسر وسهولة، العلاقات المكانية، السرعة الإدراكية، الاستدلال أو التفكير المنطقى. ويقتصر الاختبار الأخير على الصفوف من الرابع إلى نهاية المرحلة الثانوية. ويستغرق تطبيق الاختبار حوالى ساعة ونصف. وتُقرر الدرجات الخام لكل اختبار فرعى وللبطارية ككل، ثم تُحولّ إلى نسبة ذكاء انحرافية، ورتب مئينية.

(5) **اختبار كاتل للذكاء غير المتحيز ثقافياً: يهدف هذا الاختبار لقياس الذكاء السائل Fluid Intelligence، أي العامل العام (g) الذى سبق إيضاحه، ويتمثل في السلوك العقلي المتكيف في مواقف غير مألوفة، بحيث لا يسترشد الفرد في استجابته لها بمهارات تعلم سابقة. ومن أمثلة ذلك: مادة لا معنى لها، ومواد غير مألوفة، ومواد شائعة الاستخدام. وتُعرض هذه المثيرات في أشكال تتطلب تسلسل معين، أو تصنيف، أو مصفوفات، أو تحديد مواضع مناسبة لأشكال معطاة.**

تعقيب على اختبارات الذكاء العام:

يتضح مما سبق أنه توجد كثير من الاختبارات الجيدة التي تقيس القدرة العقلية العامة تناسب متطلبات المواقف التطبيقية المختلفة. وبعض هذه الاختبارات يطبق بطريقة فردية مثل: مقياس ستانفورد – بينيه، ومقاييس ويكسلر، والبعض الآخر يطبق بطريقة جماعية، ويوجد عدد كبير من هذه الاختبارات يتراوح بين تلك التي تناسب أطفال ما قبل المدرسة، وتلك التي تناسب الراشد المتفوق. وبعض هذه الاختبارات يكون على شكل بطاريات تشتمل على اختبارات فرعية تقيس كل من الجوانب اللفظية، والجوانب غير اللفظية، وينتج عنها درجات لكل من هذه الجوانب، وكذلك درجة كلية للذكاء العام. أما البعض الآخر فإنه لا يعتمد على اللغة، وإنما على الصور والأشكال والمصفوفات، وبعضها أيضاً يكون متحرراً من الثقافة بدرجة ما.

ويعتمد تفسير معظم هذه الاختبارات على الدرجات المعيارية المحولة، مثل: نسب الذكاء الانحرافية، والتساعيات المعيارية، وكذلك على المئينيات، وغيرها من الدرجات المحوّلة. ومن الجدير بالذكر أن بعض هذه الاختبارات تم تعريبه وتقنينه على بيئات عربية.

الفصل السابع

اختبارات الاستعدادات الخاصة

- مقدمة
- أنواع اختبارات الاستعدادات الخاصة
- استخدامات اختبارات الاستعدادات الخاصة
- (أولاً) بطاريات الاستعدادات الخاصة المتعددة
- اختبارات الاستعدادات الفارقة
- بطارية اختبارت الاستعداد العام
- (ثانياً) اختبارات استعدادات وقدرات نوعية
- اختبارات الاستعدادات الميكانيكية
- اختبارات الاستعداد الكتابي
- اختبارات الاستعدادات الموسيقية والفنية

مقدمة:

أوضحنا في الفصلين السابقين كيفية قياس التحصيل والذكاء، وذكرنا أن اختبارات التحصيل تهدف لقياس الأداء الحالي للفرد أو نواتج تعلمه في مدة دراسية معينة، بينما تهدف اختبارات الذكاء لقياس القدرة العقلية العامة للفرد، والتنبؤ بتحصيله الأكاديمى باستخدام درجة كلية واحدة في الاختبار. غير إن بعض اختبارات الذكاء كما رأينا في الفصل السابق ينتج عنه أكثر من درجة واحدة للفرد اعتماداً على نظريات الذكاء متعدد العومل.

فاختبارات الذكاء العام تقيس مجموعة متنوعة من القدرات، لذلك يتبين أن فائدتها في التنبؤ بمختلف المحكات ليست كبيرة. ويرى كرونباك (Cronbach, 1970) أن الاختبار الأقل اتساعاً في محتواه يمكن الإفادة منه بدرجة أفضل في التنبؤ بمحكات محدودة.

لذلك اقترح بعض علماء القياس أن اختبارات الذكاء العام ينبغي أن تطبق إلى جانب اختبارات الاستعدادات أو القدرات الكامنة، والتنبؤ بمدى استفادة الفرد من تدريب في مجال مهارى معين. ولعل هذ أيضاً هو اختلاف اختبارات الاستعدادات عن اختبارات الذكاء العام، حيث إن الهدف من اختبارات الاستعدادات التنبؤ بالنجاح في مجال أكثر محدودية مما يهدف إليه التنبؤ من درجات اختبارات الذكاء.

فاختبارات الاستعدادات الخاصة تقيس مهارات نوعية أو معارف تتعلق بمجال معين، وهذه المهارات والمعارف تعد ضرورية للأداء المستقبلى في هذا المجال. لذلك فإن اختبار الاستعداد الميكانيكي مثلاً ربما يشتمل على مهام تجميع أشياء مفككة كالجرس الكهربائي، أو جزء من أجزاء جهاز أو آلة، أو على مهام تتطلب التمييز بين مثيرات مسجلة تتباين في درجة الصوت، وعلوه، وسرعة النغمة، والإيقاع، وغير ذلك. وتعتمد هذه المهارات على تحليل العمل المطلوب Job Analysis لتحديد نوعيات الأداء الناجح في مهنة أو مجال معين.

ومن ثم يتم بناء اختبار الإستعدادات لقياس هذه المهارات أو القدرات المحددة بموضوعية ودقة.

ويتناول هذا الفصل أنواع اختبارات الاستعدادات الخاصة، واستخداماتها، وبعض بطاريات اختبارات الاستعدادات المتعددة، وأمثلة لبعض اختبارات استعدادات وقدرات نوعية تتضمن اختبارات الاستعدادات الميكانيكية، واختبارات الاستعداد الكتابي، واختبارات الاستعدادات الموسيقية والفنية.

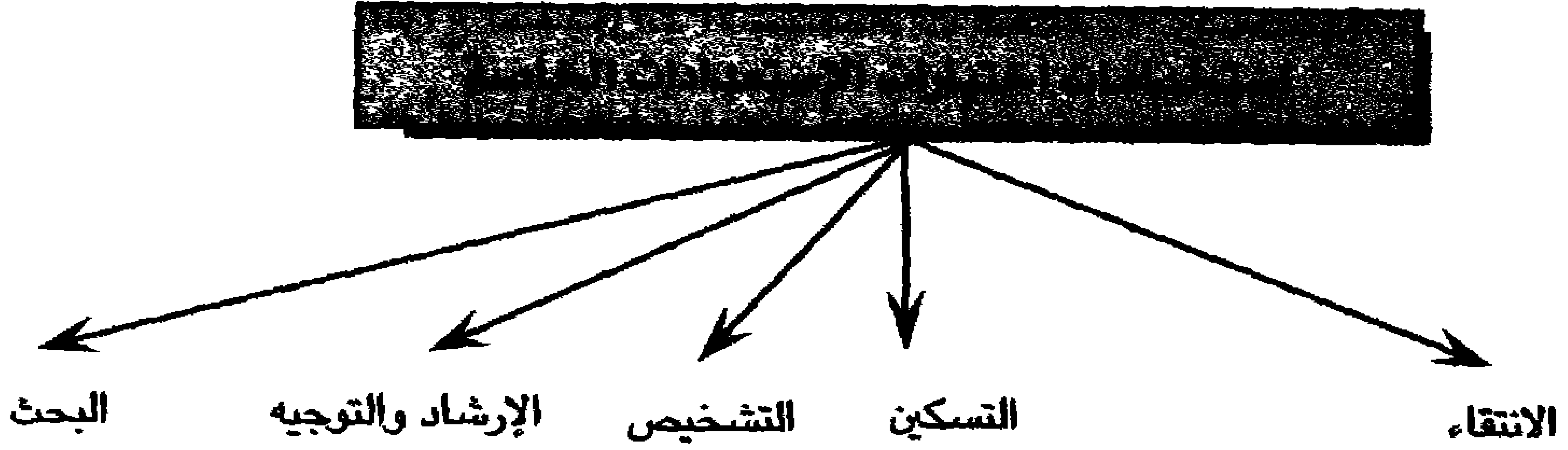
أنواع اختبارات الاستعدادات الخاصة :

تختلف اختبارات الاستعدادات الخاصة في درجة خصوصية أو نوعية القدرات أو المهارات التي تقيسها وتتطلبها مهن أو أعمال أو أنشطة معينة. وبعض هذه الاختبارات يُستخدم في التنبؤ بالنجاح أو الفشل في دراسات معينة، أو في تعلم مجال دراسي تخصصي، كما يُستخدم في التنبؤ بالأداء في مهنة أو عمل معين. ففي الحالة الأولى نحن نهتم بالاستعداد الدراسيScholastic Aptitude، وفي الحالة الثانية نهتم بالاستعداد المهني Vocational Aptitude. وبعض اختبارات الاستعدادات يمكن أن يحقق هدفين معاً، حيث يقيس الذكاء العام، وكذلك الاستعداد الدراسي. ويوجد عدد كبير من اختبارات الاستعداد الدراسي في مجالات متنوعة، مثل:الفنون الحرة، والطب، والتمريض، وغير ذلك. كما يوجد أيضاً عدد كبير من اختبارات الاستعداد المهني الميكانيكي، واليدوي، والكتابي أو التجاري، والموسيقي، والفنى، وغيرها.

وتوجد أيضاً بطاريات واختبارات استعدادات متعددة Multiple Aptitude Tests Batteries تُستخدم في أغراض الإرشاد التربوي والتوجيه المهني، حيث إن كلاً منها يشتمل على عدة اختبارات تقيس استعدادات خاصة مختلفة. وبعض هذه الاختبارات يُضم معاً في بطاريات تزودنا بمعلومات تتعلق بكل من الاستعدادات الدراسية والمهنية.

استخدامات اختبارات الاستعدادات الخاصة:

تُستخدم اختبارت الاستعدادات الخاصة بكثرة في المؤسسات التعليمية والمهنية لأغراض متعددة يوضحها الشكل التخطيطي (7 - 1) التالي:



شكل (7 - 1): يوضح اغراض استخدامات اختبارات الاستعدادات الخاصة

وفيما يلي توضيحاً لكل من هذه الاستخدامات:

أغراض الانتقاء: وتُعد من أهم استخدامات اختبارات الاستعدادات. غير إن استخدام نتائج هذه الاختبارات في أغراض الانتقاء Selection لا يضمن اتخاذ قرارات صائبة فيما يتعلق بقبول الأفراد أو رفضهم في جميع الحالات، بل نتوقع بعض الخطأ في هذه القرارات.

لذلك يهتم القائمون ببناء اختبار استعداد معين التحقق من الصدق المرتبط بمحك للاختبار، بحيث تكون درجاته مرتبطة بمحك معين ارتباطاً مرتفعاً.

اغراض التسكين: وهنا يُفترض أن الأفراد قد تم انتقاؤهم في مؤسسة تربوية أو مهنية معينة، ويكون مطلوباً تسكينهم في مجال دراسي، أو عمل معين مناسباً لهم ومتفقاً مع ميولهم. وكثير من المشكلات المتعلقة باتخاذ قرارات الانتقاء تتواجد أيضاً عند استخدام اختبارات الاستعدادات في أغراض التسكين Placement .

فالبعض يرى أن اتخاذ قرارات التسكين تكون لصالح المؤسسة، لذلك ربما يتم تسكين الأفراد الذين يتبين أن لديهم الاستعدادات الجيدة في عمل معين، مما يجعل هؤلاء الأفراد يضطرون لقبول هذا العمل على الرغم من أنهم ربما يرون أنهم لا يميلون إليه، أو أنه غير مناسب لهم. وهذه تعد إحدى مشكلات استخدام اختبارات الاستعدادات في أغراض التسكين.

لذلك ينبغي أن تكون قرارات الانتقاء مسئولية مشتركة بين المؤسسة والأفراد المتقدمين للدراسة أو العمل.

أغراض التشخيص: تفيد اختبارات الاستعدادات فائدة كبيرة في تشخيص مشكلات الطلاب. إذ تطبق عادة اختبارات التهيؤ Readiness Tests في المرحلة الابتدائية للتعرُّف على التلاميذ الذين يُحتمل مواجهتهم بعض الصعوبات أثناء تعلم مواد دراسة معينة، مثل : القراءة أو الحساب، كما سبق أن أوضحنا في الفصل الخامس.

وكذلك في المدرسة الثانوية ربما تُطبق اختبارات استعدادات في اللغات الأجنبية لتحديد الطلاب الذين لديهم المهارات اللازمة لتعلم لغات أجنبية معينة. وتحليل أخطائهم في هذه الاختبارات يساعد في تصميم الإجراءات العلاجية المناسبة.

أغراض الإرشاد والتوجيه: تفيد أيضاً اختبارات الاستعدادات في الإرشاد والتوجيه التربوي، والمهني، والشخصي. وتُستخدم عادة بطاريات اختبارات الاستعدادات متعددة العوامل Multifactor Aptitude Tests في الإرشاد والتربوي Educational counseling لتعريف الطلاب بقدراتهم المتمايزة. وعلى الرغم من أن قياس هذه القدرات المتمايزة لا يؤدى بالضرورة إلى تحسين التنبؤ، إلا أنه يجعل الطالب مدركاً لجوانب أو مجالات تميزه. ولكن ينبغى الحيطة من إساءة استخدام نتائج بطاريات اختبارات الاستعدادات في هذا الغرض، إذ لا يجوز للمرشد التربوي تفسير درجات منخفضة في اختبار معين بطريقة تؤثر سلبياً في مفهوم الذات لدى الطالب، وإنما ينبغى توظيف الدرجات المرتفعة في الاختبارات الأخرى في حفز دافعية الطالب وإشعاره بقدرته. كما ينبغى التأكيد على أن درجات اختبارات الاستعدادات بمفردها لا تعد كافية للإرشاد والتوجيه التربوي والمهني، وإنما يجب أيضاً الأخذ بعين الاعتبار أداء الفرد السابق، وميوله، والعوامل الموقفية المختلفة.

أغراض البحث: يمكن استخدام اختبارات الاستعدادات في كثير من البحوث النفسية والتربوية. فدرجات هذه الاختبارات تُستخدم عادة كمتغيرات مستقلة في تصميم البحوث، على الرغم من إمكانية استخدام هذه الدرجات كمتغيرات تابعة، كما هو الحال في دراسة تأثيرات البيئة في الذكاء. كما أن البحوث التي تُجري حول طبيعة الذكاء والاستعدادات الخاصة تؤدى إلى تطوير الاختبارات ذاتها، وتُقدم بيانات مهمة تفيد في إعادة النظر في بعض نظريات التكوين العقلي، وفهم النمو المعرفي.

(أولاً) بطاريات الاستعدادت الخاصة المتعددة :

تعد بطاريات الاستعدادات الخاصة المتعددة Multiaptitude Batteries من المقاييس التي استندت في بنائها على أسلوب التحليل العاملى الذي سبق أن أشرنا إليه في مستهل الفصل السابق من خلال تناولنا نظريات بنية الذكاء. فقد اعتمد ثيرستون، وجيلفورد، وغيرهما على هذا الأسلوب في التوصل إلى الأبعاد أو العوامل التي أُطلق عليها القدرات العقلية. فمحاولات عزل وتحديد عوامل نوعية للذكاء العام أثّرت في جهود علماء القياس المتعلقة بتحديد الاستعدادات الخاصة وقياسها.

فقد حاول هؤلاء العلماء بناء بطاريات من الاختبارات لقياس قدرات أو سمات متعددة تتطلبها أنواع مختلفة من الأعمال أو الأنشطة، فإذا ارتبطت درجات الاختبارات التي تشتمل عليها البطارية بمهن معينة أو مجموعات من المهن، فإنه يمكن تقدير أوزان لهذه الدرجات وفقاً لأهميتها في هذه المهن (Aiken, 1976 & Noll, scannel, 1972) .

لذلك تسمى هذه البطاريات أحياناً بطاريات استعدادات فارقة، أو بطاريات استعدادات عاملية " .

والاختبارات التي تشتمل عليها هذه البطاريات تقيس عادة عدة قدرات مستقة، ومعاييرها تسمح بإجراء مقارنات بين قدرات الفرد نفسه، لذلك تعتمد معايير اختبارات البطارية جميعها على نفس عينة التقنين.

وتفيد هذه البطاريات في الإرشاد والتوجيه والمهني، وانتقاء الأفراد وتصنيفهم، حيث يكون هناك اهتمام بقياس مهارات الفرد ومعارفه في مجالات متعددة. وتعتمد معظم هذه البطاريات عادة على مفردات اختبارية تتطلب الورقة والقلم، لذلك يمكن تطبيقها على مجموعة كبيرة من الأفراد في وقت واحد. والمعلومات المستمدة من هذه البطاريات تفيد الطالب في معرفة جوانب قوته وضعفه، ويمكن أن يُستند إليها في اتخاذ قرارات تربوية ومهنية تتعلق بمستقبله. غير إنه لا يُنصح عادة باستخدام بطاريات الاستعدادات المتعددة قبل وصول الطالب إلى المرحلة التعليمية المتوسطة نظراً لأن قدرات الأطفال تميل إلى العمومية في المرحلة الابتدائية .

ويوجد العديد من بطاريات الاستعدادات المتعددة، غير أن أقدمها وأكثرها شيوعاً واستخداماً البطاريتين التاليتين :

(1) اختبارات الاستعدادات الفارقة (Differential Aptitude Tests (DAT :

أعد هذه البطارية من الاختبارات بينيت، وسيشور، وويزمان (Bennett, Seashore & Wesman, 1974) وصدرت طبعتها الخامسة عام 1973، وأُجريت عليها بعض التعديلات عام 1982. والغرض الأساسي من هذه البطارية تقديم معلومات تتعلق ببروفيل قدرات متعددة لاستخدامه في التوجيه والإرشاد التربوي للطلاب في المرحلتين المتوسطة والثانوية لكي يتخذ كل منهم في ضوء ذلك قرارات تتعلق باختياراته المهنية المستقبلية.

وتشتمل البطارية على ثمانية اختبارات فرعية يُطبق كل منها على حدة، وهى كالتالي:

(1) الاستدلال اللغوي (تكملة جمل ناقصة باستخدام كلمات معطاة).

(2) القدرة العددية (حساب، وعمليات حسابية) .

(3) الاستدلال المجرد (إكمال متسلسلات شكلية) .

(4) العلاقات المكانية (تكوين شكل معين من أنماط شكلية معطاة).

(5) الاستدلال الميكانيكي (إجابة أسئلة تتعلق بمبادىء ميكانيكية ممثلة بصور معطاة) .

(6) سرعة النسخ ودقته (التعرف على مجموعات معينة مركبة من حروف وأرقام) .

(7) استخدام اللغة (التهجى، والنحو، وعلامات الترقيم) .

وإجابة بعض مفردات هذه الاختبارات تتطلب الاختيار من متعدد. ومجموع درجات اختبار الاستدلال اللغوي، واختبار القدرة العددية يُعطي مؤشراً للاستعداد الدراسي. ويستغرق تطبيق هذه البطارية ثلاث ساعات تقريباً، وهي تعتمد على القوة . وتقرر النتائج للطلاب في شكل بروفيل كالتالي:

المئينيات ذكور	المعايير المئينية (5 10 20 25 30 40 50 60 70 75 80 90 90 19)	المئينيات إناث
استدلال اللغوي 70		70
الدرجة الخام 25		
القدرة العددية 75		75
الدرجة الخام 25		
معدل اللغوي+ القدرة العددية75		70
الدرجة الخام 51		
الاستدلال المجردة 60		55
الدرجة الخام 31		
العلاقات المكانية 60		65
الدرجة الخام 28		
الاستدلال الميكانيكي الدرجة 45		75
الدرجة الخام 46		
سرعة النسخ ودقته 25		40
الدرجة الخام 36		
استخدام اللغة 65		45
الدرجة الخام 24		

شكل (7-2): يوضح بروفيل أحد طلاب الصف الأول الثانوي في بطارية اختبارات الاستعدادات الفارقة

ويتضح من شكل (7 - 2) أن الدرجات الخام لكل اختبار فرعي مدونة في العمود الأول، والرتب المئينية المناظرة لها في العمود الثاني. وهذه الرتب توضح للطالب رتبته بالنسبة لأقرانه من نفس النوع ونفس الصف المدرسى على المستوى القومى.

بينما دُونت الرتب المئينية التي تقارن أداء الطالب بأداء أقرانه في نفس الصف المدرسي ولكن من النوع الآخر في العمود الأخير .

وقد رُسم شريط من العلامات (X) يمثل الرتبة المئينية للطلاب من نفس النوع لكل اختبار فرعي، بحيث أن الرتبة المئينية لكل من هذه الاختبارات المدونة في العمود الثانى تقع في منتصف الشريط (X). والشريط المناظر لكل اختبار فرعي يدل على أن ثبات الدرجات ليس تاماً، وطول كل شريط يعكس الخطأ المعيارى للقياس.

ويكون مستوى قدرة الطالب في أي مجالين متساوياً إذا كان هناك تداخل في نهايات الشريطين الممثلين لهذين المجالين. أما إذا لم يكن هناك تداخل، فإن قدرة الطالب في المجال الذي حصل فيه على درجة أعلى يُحتمل أن تكون قدرته فيه أعلى. فإذا نظرنا إلى شكل (7-2) نجد أنه يمكن تفسير درجة الطالب في القدرة العددية على أنها أعلى من قدرته على استخدام اللغة، وذلك لعدم تداخل الشريطين.

وإذا كان هناك تداخل ولكن أقل من نصف طول الشريطين، فإن الفرق بين الدرجتين ربما يكون له دلالة، ولكننا نحتاج في هذه الحالة إلى مزيد من المعلومات عن الطالب قبل اتخاذ قرار بشأنه .

وقد أُعدت معايير الرتب المئينية، والتساعيات المعيارية لكل اختبار فرعي، ولمجموع الدرجات الخام لاختباري الاستعداد اللغوي، والقدرة العددية. كما أُعدت معايير منفصلة للذكور، والإناث، والصفوف المدرسية (8-12).

وتم تقدير ثبات درجات كل من هذه الاختبارات الفرعية على حدة – فيما عدا اختبار سرعة النسخ ودقته – بطريقة التجزئة النصفية. ويرجع ذلك إلى أن هذا الاختبار يعتمد على السرعة في الإجابة، لذلك تم تقدير ثبات درجاته باستخدام طريقة الصيغ المتكافئة.

وقد تراوحت قيم معامل ثبات اختبارات البطارية بين (0.97،0.87). للذكور، وبين (0.96،0.83) للإناث، وقد أُجرى العديد من دراسات الصدق التنبؤى لهذه البطارية من

الاختبارات، وذلك بإيجاد الارتباطات بين درجات الاختبارات الفرعية وتقديرات الطلاب في مختلف المواد الدراسية، ودرجاتهم في اختبارات الذكاء، وكذلك درجاتهم في بطاريات أخرى للاستعدادات، واختبارات تحصيل موضوعية، وبيانات مهنية. وقد بينت النتائج بعامة أنه توجد علاقة تمايزية بين هذه الاختبارات والتحصيل في مختلف المواد الدراسية، ومستويات القدرة العقلية، والتصنيفات المهنية.

(2) بطارية اختبارات الاستعداد العام: General Aptitude Test Battery (GATB):

أعد هذه البطارية مكتب خدمات التوظيف بالولايات المتحدة الأمريكية، واعتمد بناؤها على تحليل مختلف الأعمال، ونتائج التحليل العاملي لعدد كبير من الاختبارات المختلفة المتعلقة بأعمال معينة، وبلغ عددها (59) اختباراً.

واختير فقط (10) عوامل قدرات مختلفة لكي تقيسها البطارية. وتشتمل البطارية في طبعتها الحديثة على (12) اختباراً فرعياً، ينتج عنها درجات لتسعة عوامل مختلفة، وأحد هذه العوامل هو القدرة العقلية العامة (G) الذي ينتج من درجات ثلاثة اختبارات فرعية هي: المفردات اللغوية، والاستدلال الحسابي، والتصور المكاني في ثلاثة أبعاد .

وفيما يلي توضيحاً للاختبارات الفرعية اتى تشتمل عليها البطارية :

(1) الذكاء : وهو مركب من الاختبارات الفرعية الثلاثة المشار إليها .

(2) القدرة اللفظية : مفردات لغوية، استخدام المترادفات، والمناقضات في المعني .

(3) القدرة العددية : إجراء عمليات حسابية، واستدلال حسابي.

(4) القدرة المكانية : القدرة على تصور أشكال في الفراغ ثلاثي الأبعاد، ولكنها عُرضت في بعُدين .

(5) إدراك صيغ وأنماط بصرية : مزاوجة رسوم لأدوات وأشكال هندسية .

(6) الإدراك الكتابي : مزاوجة أسماء .

(7) التآزر الحركي : القدرة على وضع علامات معينة داخل مجموعة من المربعات .

(8) مهارة الأصابع : القدرة على تجميع وفك مسامير البرشام، والحلقات المعدنية .

(9) المهارة اليدوية : القدرة على النقل والتركيب في لوحة مثقبة .

ويمقارنة هذه الاختبارات بالاختبارات الفرعية التي اشتملت عليها بطارية الاستعدادات الفارقة نلاحظ أن البطارية الأخيرة تحتوى على اختبارات للفهم الميكانيكي و اللغة، وهو ما لم تشتمل عليه بطارية اختبارات الاستعداد العام. في حين أن هذه البطارية اشتملت على اختبارات حركية لم تشتمل عليها بطارية الاستعدادات الفارقة. لذلك فإن بطارية اختبارات الاستعداد العام موجهة للعمل بدرجة أكبر من توجهها للأداء الأكاديمي. وقد وُجد أن الاختبارات الحركية الثلاثة التي تشتمل عليها هذه البطارية ترتبط فيما بينها ارتباطاً مرتفعاً نسبياً، ولكنها في الواقع لا ترتبط بالاختبارات الفرعية الأخرى.

كما وُجدت ارتباطات مرتفعة إلى حد ما بين العوامل المتناظرة التي تقيسها كل من البطاريتين. والفرق الرئيس بينهما هو أن اختبارات البطارية الأولى (DAT) تعد في معظمها اختبارات قوة Power Tests ، أما اختبارات البطارية الثانية (GATB) فتُعد من الاختبارات التي تعتمد إلى حد كبير على السرعة Speed Tests . لذلك تفيد هذه البطارية في التنبو بمحكات تتطلب التناسق الحركى البصري .

والدرجات الخام للعوامل التسعة المستمدة من الاختبارات الفرعية تُحوّل إلى رتب مئينية أو درجات معيارية متوسطها (100)، وانحرافها المعياري (20). ويمكن مقارنة الدرجات المعيارية للفرد بتلك المستمدة من (36) نمطاً من أنماط الاستعدادات المهنية الناتجة عن تحليل درجات الأفراد في هذه البطارية الذين ينتمون إلى أكثر من (800) مهنة أو عمل. وحُددت درجات قطع Cutting Scores لكل اختبار ينبغي أن يصل إليها الفرد للأداء الفاعل في المهنة المعينة.

ويستغرق تطبيق هذه البطارية قرابة الساعتين ونصف، وتناسب طلاب نهاية المرحلة الثانوية، وكذلك الراشدين. وقد تراوحت قيم معاملات ثبات درجات الاختبارات الفرعية باستخدام طريقتي التجزئة النصفية، والصيغ المتكافئة بين (0.80 ، 0.90)، والخطأ المعيارى للقياس (7) نقاط بالنسبة للدرجات المعيارية. كما أن صدق الدرجات التسع، وأنماط القدرات المهنية التي عددها (36) نمطاً في مقابل محكات مهنية وأكاديمية تراوح بين (صفر، 0.90) .

وفى ضوء هذه المعلومات يمكن اعتبار هذه البطارية (GATB) من أفضل البطاريات التي تُستخدم في الإرشاد المهني، وانتقاء الأفراد، في حين أن بطارية الاستعدادات الفارقة (DAT) يبدو أنها تناسب بدرجة أفضل أغراض الإرشاد التربوي.

ومن الجدير بالذكر أن هناك بطاريات أخرى مثل : بطارية فلاناجان لتصنيف الاستعدادات Flanagan Aptitude Classificatian (FACT)، وبطارية فلاناجان الصناعية Flanagan Industrial Tests (FIT), ، وبطارية جيلفورد – زيمرمان لمسح الاستعدادات Guilford - Zimmerman Aptitude Survey، وجميعها تقيس جوانب القوة والضعف في الاستعدادات المختلفة لدى الأفراد.

(ثانيا) اختبارات استعدادات وقدرات نوعية :

تقيس بطاريات الاستعدادات المتعددة التي أوضحنا بعضها مجموعة من القدرات في بطارية اختبارات واحدة متكاملة. غير إنه يمكن أيضاً بناء اختبارات منفصلة لقياس استعدادات أو قدرات نوعية Specific Aptitudes . لذلك سوف أوضح فيما يلي بعض أهم هذه الاختبارات التي يُستخدم بعضها في المواقف الأكاديمية، وبعض آخر في المواقف المهنية. وفي أي الحالات يهتم مصمم الاختبار بقياس قدرة نوعية محددة قياساً أكثر عمقاً بدلاً من اهتمامه بتقديم بروفيل لعدة قدرات.

وهذا يجعل الاختبار أكثر صدقاً لتعلقه بغرض أو موقف معين. غير إنه نظراً لأن اختبارات الاستعدادات النوعية المختلفة تستند معاييرها إلى مجتمعات مختلفة من الأفراد، فإن البيانات التي تُستخدم في حساب المعايير لا تكون متقاربة، مما يؤدي إلى صعوبات في تفسير الدرجات وبخاصة إذا أُجريت مقارنات فيما بينها .

(1) اختبارات الاستعدادات الميكانيكية Mechanical Aptitude Tests:

يرى بعض خبراء القياس أن الاستعداد الميكانيكي يُعد سمة أحادية البُعد يختلف مقدارها من فرد إلى آخر، بينما يرى بعض آخر أنه سمة متعددة الأبعاد، أى تشتمل على عدد من القدرات المستقلة نسبياً. ولعل هذا المدخل الذى يمايز بين مكونات الاستعداد الميكانيكي هو الذي حظي بتأييد أكبر. وهذا يعني أن قياس هذا الاستعداد بدرجة أفضل يتطلب تطبيق عدة اختبارات أو اختبارات فرعية وليس اختباراً واحداً.

وبعض هذه الاختبارات يقيس أداءات أو مهارات يدوية أو نفسية حركية معينة، وبعض آخر يقيس معلومات ميكانيكية تعتمد على الورقة والقلم. وفيما يلي بعض اختبارات الاستعدادات الميكانيكية :

اختبار التجميع الميكانيكي Mechanical Assembly Test :

تتطلب هذه الاختبارت تجميع الأجزاء، وتعتمد درجة الفرد على عدد الأجزاء التي يمكنه تجميعها في زمن محدد. ولعل اختبار مينسوتا للتجميع Minnesota Assembly Test يعد أحد أهم هذه الاختبارات. ويشتمل على (33) شيئاً أو شكلاً ميكانيكياً مفككاً ينبغي على الفرد تجميعها في زمن محدد. غير أن اختبارات التجميع الميكانيكي كلفتها عالية وصدقها منخفض حيث، لا ترتبط درجات اختبار مينسوتا بتقديرات الطلاب.

ونظراً لأن معايير هذا الاختبار مستمدة من طلاب المدارس الثانوية، فإن المعلومات المتعلقة بقدرة الاختبار على التنبؤ بالنجاح في مهنة معينة غير متوافرة .

اختبار المعلومات الميكانيكية Mechanical Information Test:

يقيس هذا الاختبار معلومات الفرد أو خبرته الميكانيكية، حيث يمكن باستخدامه انتقاء القائمين بتدريب الأفراد، ويُفترض أن هؤلاء المدربين لديهم الاستعداد الميكانيكي.

اختبار أوروك للاستعداد الميكانيكي O ' Rourke Mechanical Aptiltude Test:

ويشتمل على مفردات تتطلب المزاوجة بين الصور والأدوات، وهذه المفردات تمثل الجزء الأول من الاختبار، أما الجزء الثاني فيشتمل على مفردات اختيار من متعدد تقيس المعلومات المتعلقة بالأدوات واستخداماتها. ومفردات كل من الجزئين تحتوي على أسئلة متنوعة ومتعددة تقيس المعلومات في مجالات النجارة، والكهرباء، والطلاء، والطباعة، وغيرها من المهن. غير إن البيانات المتوافرة حول صدق الاختبار وثبات درجاته محدودة .

اختبار الاستدلال الميكانيكي Mechanical Reasoning Test :

يُعد اختبار بينيت للفهم الميكانيكي Bennett Mechanical Comprehension Test (BMCT) واحداً من أفضل اختبارات الاستدلال الميكانيكي. ويشتمل على أسئلة تحتوى على صور لأشياء مألوفة بينها علاقة ميكانيكية معينة، وعلى الفرد أن يتوصل إلى استدلالات صحيحة عن هذه العلاقات. ومن أمثلة ذلك: صورة لغرفتين تحتويان على عدد مختلف من قطع الأثاث، ويُطلب من الفرد الاستدلال على الغرفة التي يكون صدى الصوت فيها أكبر.

وقد أُعدت لهذا الاختبار معايير، وكذلك أُعدت صيغة أخرى للاختبار تناسب الإناث.

(2) اختبارات الاستعداد الكتابي Clerical Aptitude Tests:

يُعد الاستعداد الكتابي أيضاً من الاستعدادات متعددة الأبعاد، وعلى الرغم من أن المهارة اليدوية، وسرعة إدراك التماثل والاختلاف بين الأشياء يُعد ضرورياً للعمل الكتابي، إلا أن القدرات اللغوية والرياضية التي تعد من مظاهر الذكاء العام، تكون مهمة أيضاً في العمل الكتابي. لذلك فإن كثيراً من اختبارات الاستعداد الكتابي تشتمل على مفردات مشابهة لتلك التي تحتويها اختبارات الذكاء العام، إضافة إلى مفردات تقيس السرعة الإدراكية، والدقة وكثير من اختبارات الاستعداد الكتابي الأكثر شمولية تحاول قياس استعدادات متنوعة، مثل : المهارات اللفظية، وملأ فراغات بحروف، وقواعد لغوية، وتهجٍّي، وجميعها مهارات مهمة للنجاح في العمل الكتابي. ومن الملاحظ أن بطارية اختبارات الاستعدادات الفارقة (DAT) التي أوضحناها تحتوى على اختبارات فرعية تتعلق بالاستعداد الكتابي .

اختبار مينسوتا للاستعداد الكتابي : Minnesota Clerical Test:

يعد هذا الاختبار من اختبارات الاستعداد الكتابي العام، وأعده أندرو Andrew، وبترسون Paterson. ويُستخدم بكثرة في انتقاء العاملين في المجالات الكتابية، والمدققين، وغيرهم ممن يلتحقون بوظائف تتطلب السرعة الإدراكية، والتعامل مع الرموز بدقة. ويشتمل الاختبار على جزئين هما: المقارنة بين الأعداد، والمقارنة بين الأسماء، حيث يتحقق المختبر من الأزواج المتطابقة في قائمة تشتمل على (200) زوج من الأعداد، (200) زوج من الأسماء على الترتيب. وكل من الجزئين له زمن محدد، ويتم تصحيح الدرجات من أثر التخمين في الإجابة.

وتشتمل معايير الاختبار على المئينيات لكل من الذكور والإناث، والفرق الدراسية (7- 12)، ولمجموعات من العاملين في المجالات الكتابية، والمتقدمين لشغل هذه الوظائف .

ومعامل ثبات درجات الاختبار بطريقة التجزئة النصفية تتراوح بين (0.70، 0.80) وترتبط هذه الدرجات بتقديرات المعلمين والمشرفين على الأعمال الكتابية ارتباطاً متوسطاً .

وتوجد اختبارات أخرى تقيس الاستعداد الكتابي العام وتتناول مفرداتها: التهجِّي، الترتيب الأبجدي، الجمل غير المرتبة، العمليات الحسابية البسيطة.

وتوجد أيضاً اختبارات استعدادات تتعلق بمجالات خاصة، مثل: الاختزال، وبرمجة الحاسوب، وغير ذلك من المجالات الكتابية النوعية .

(3) اختبارات الاستعدادات الموسيقية والفنية Musical and Artistic Tests:

توجد اختبارات استعدادات موسيقية وفنية متعددة، غير إنها محدودة الاستخدام بوجه عام. ولعل أكثر اختبارات الاستعدادات الموسيقية شيوعاً اختبار سيشور للموهبة الموسيقية Seashore Measures of Mental Tests.

وتُقدَّم مفردات هذه الاختبارات مُسجلة على شرائط، وتشتمل على ستة اختبارات هي: تمييز درجة النغم، وتمييز علو الصوت، وتمييز الزمن، والحكم على الزمن، والحكم على الإيقاع، وذاكرة النغم. ويعتبر سيشور وزملاؤه أن هذه القدرات أساسية في تنمية الكفاءة في الموسيقى. وهذه الاختبارات تناسب الطلاب من الصف الرابع إلى الرشد ويستغرق تطبيقها ساعة واحدة تقريباً بمعدل (10) دقائق لكل اختبار.

وقد وُجد أن معامل ثبات درجاتها بطريقة التجزئة النصفية يتراوح بين (55.0 ، 0.90). وتم حساب المعايير المئينية للصفوف (4 - 5)، (6 - 8) ، (9 - 16). وبينت نتائج البحوث أن درجات هذه الاختبارات ترتبط بمحكات التدريب الموسيقي ارتباطاً تتراوح قيمته بين (0.30 ، 0.40)، ولكن البيانات المتعلقة بصدق الاختبارات غير كافية. ويُعاب على هذه الاختبارات أنها جزئية، أي أن كلاً منها يقيس جانباً محدوداً من الموهبة الموسيقية.

أما اختبارات الاستعدادات الفنية فإنها تنقسم إلى نوعين هما:

الحكم الفني Aesthetic Judgment، والإنتاج الفني Aesthetic Production. ويُعد الحكم الفني ضرورياً للنقد الفنى، أما الإنتاج الفنى فإنه يتعلق بالعمل الفني، ويتطلب الإنتاج الفنى أيضا حكماً فنياً، ولكن العكس لا يكون ضرورياً.

اختبار ماير للحكم الفني Meier Art Judgment Test

يُعد أكثر اختبارات الاستعدادات الفنية شيوعاً، ويشتمل على (100) زوج من اللوحات الفنية الشهيرة، وإحدى لوحات كل زوج تُعد أصلية، أما الأخرى فإنها تُعد نسخة معدلّة من الأولى. ويُطلب من الفرد تحديد أوجه الاختلاف بينهما، وأيهما يفضل. ويمكن باستخدام هذا الاختبار التمييز بين مجموعات الأفراد الذين يختلفون اختلافات محددة في خبرتهم الفنية .

قائمة هورن للاستعداد الفني Horn Art Aptitude Inventory:

تُعد مقياساً للإنتاج الفني. وتشتمل على جزئين : أحدهما يتطلب رسم (20) شيئاً مختلفاً ولكنه مألوف، مثل: شجرة، أو كتاب، أو ملعقة، في زمن يتراوح بين (10،3) ثواني. أما الآخر فيتطلب التخيل، حيث يشتمل على (12) مستطيلاً يتضمن كل منها خطوطاً متعددة يستند إليها الفرد في رسم صورة مُمثِّلة. وتتميز هذه القائمة بصدق مرتفع، كما أن قيمة معامل ارتباط درجاتها بمتوسط تقديرات معلمى التربية الفنية بلغت (0.53).

ومن هذا يتضح تنوع الاستعدادات الخاصة، وتعدد الاختبارات التي تقيسها والتي تختلف في مدى اتساع محتواها وأغراض استخداماتها. والاختبارات التي تناولها هذا الفصل تُعد بمثابة عينة من كثير من اختبارات الاستعدادات التي يعرضها الكتاب السنوى للقياس العقلي Buros ' Mental Measurement Year book.

غير إن هذه العينة اشتملت على بعض الاختبارات الأكثر شيوعاً واستخداماً في التنبؤ بالأداء الأكاديمى، والكفاءة في مجالات مهنية متنوعة .

الفصل الثامن

اختبارات ومقاييس الشخصية

- مقدمة
- مفهوم الشخصية
- أهداف قياس الشخصية
- بعض مشكلات قياس الشخصية
- استراتيجيات بناء استبيانات الشخصية
- أنواع اختبارات ومقاييس الشخصية
- (أولاً) استبيانات تستند إلى الاستراتيجية المنطقية
- (ثانياً) استبيانات تستند إلى الاستراتيجية الإمبريقية
- (ثالثاً) استبيانات تستند إلى استراتيجية التجانس
- تعقيب عام على استبيانات الشخصية
- المقاييس الإسقاطية للشخصية
- تعقيب عام على المقاييس الإسقاطية للشخصية

مقدمة:

عرضنا في الفصول الثلاثة السابقة بعض أهم الاختبارات والمقاييس التي تُستخدم في تقييم الخصائص أو السمات المعرفية لدى الأفراد. غير أن الشخصية الإنسانية تُعد تركيباً متفرداً من الخصائص المعرفية وغير المعرفية، أي الوجدانية، التي تمثل نمط السلوك المتسق والمميز للفرد. لذلك فإن قياس وتقييم الشخصية يشتمل على نطاق واسع من المتغيرات المعرفية وغير المعرفية، مثل: التحصيل، والذكاء العام، والاستعدادات الخاصة التي أوضحناها، وكذلك الميول، والاتجاهات، والقيم. كما أن هناك خصائص أخرى مميزة للفرد تتعلق بالجوانب الإنفعالية والمزاجية التي يُشار إليها عادة بأنها متغيرات شخصية Personality Variables. وهذه الجوانب لها أهمية كبيرة في فهم سلوك الأفراد والتنبؤ به . وبالطبع يصعب الفصل بين الخصائص المعرفية وغير المعرفية في الشخصية الإنسانية، إذ أنها تعد خصائص متكاملة وتؤثر في سلوك الأفراد. وتناول كل منها على حدة لا يعني أنها تؤثر تأثيرا منفرداً في السلوك، وإنما لتيسير دراستها وفهمها.

لذلك سوف أفرد هذا الفصل لتوضيح الأساليب المختلفة التي تُستخدم في قياس الشخصية الإنسانية من جوانبها الانفعالية والمزاجية.

ويتضمن ذلك التعريفات المتعددة لمفهوم الشخصية، وأهداف قياس الشخصية، وبعض مشكلاته، والاستراتيجيات المختلفة لبناء مقاييس الشخصية، وأنواع هذه المقاييس، وأمثلة لكل نوع منها متضمنة مكونات المقياس، وكيفية تفسير درجاته، واستخداماته، وميزاته وعيوبه .

مفهوم الشخصية Personality :

يُعد مفهوم الشخصية من المفاهيم المعقدة التي تتميز بالغموض، إذ لا يوجد اتفاق بين علماء النفس أنفسهم حول طبيعة الشخصية الإنسانية. فهذا المفهوم مشتق من المصطلح اللاتيني " Persona " التي تعني "القناع "، أو «الشخص المختفي بهذا القناع». وقد أُضيفت

دلالات أخرى لهذا المفهوم تدريجياً تعبّر عنها الكلمة الإنجليزية " Personage "، أي " أشخاص ذوى أهمية " . ولعل هذا يتضح من تعدد تعريفات هذا المفهوم الذي أشارت إليه الأدبيات؛ فأحد المراجع المهمة التي تناولت نظريات الشخصية عرضت ما يزيد عن عشر نظريات رئيسة، ويندرج تحتها نظريات فرعية (Hall & Lindzey, 1957) . ويمكن وضع معظم نظريات الشخصية على مُتصل يتراوح فيه التأكيد التام بين السلوك الملاحظ والعمليات الداخلية التي تنعكس في السلوك. ولعل نظريات الشخصية، التي تبنّاها علماء النفس في أمريكا تميل إلى تأكيد المنظور السلوكي، حيث يمكن تعريف الشخصية بأنها مجموع السلوك الكلي للفرد، أو أنها النزعة المميزة للفرد إلى أن يسلك بطريقة متشابهة في المواقف المتشابهة. أما النظريات التي تبنّاها علماء النفس في أوروبا، فإنها تميل إلى تأكيد الطرف الآخر للمتصل، أى العمليات الداخلية.

ويتمثل ذلك في نظريات التحليل النفسى، وبخاصة نظرية فرويد Freud. وعلى الرغم من تأكيد علماء النفس الأمريكيين على السلوك بدرجة أساسية، إلا أن المعاصرين منهم اعترفوا بوجود دوافع كامنة، أو حالات عضوية داخلية تحفز الفرد للسلوك، أو حاجات تؤدى إلى تنشيط السلوك وتوجّهه. وبالطبع يصعب في هذا الكتاب المدخلى مناقشة النظريات الرئيسة للشخصية، وإنما يمكن الإشارة إلى بعض التعريفات الشائعة للشخصية كمفهوم سيكلوجى. وربما تعتمد هذه التعريفات على ما إذا كانت تنظر إلى الشخصية على أنها كينونة Entity متمايزة عن الكينونات الأخرى ولها خصائص معينة، أم تنظر إليها على أنها وصف لكيفية اختلاف الأفراد في من درجة امتلاكهم مجموعة من السمات. والواقع أن كلاً من النظرتين تكمل إحداهما الأخرى. فقد عرَّف ألبورت Allport الشخصية بأنها التنظيم الدينامي داخل الفرد للأنظمة السيكولوجية التي تحدد تكيفه المتفرد لبيئته. ومع هذا فإنه في مراجعته للأدبيات التي تناولت معاني مفهوم الشخصية، استطاع ألبورت التمييز بين 50 معنى مختلفاً، وربما هناك المزيد من المعانى.

ويمكن تصنيف تعريفات الشخصية بعامة في خمسة أقسام منفصلة كالتالي:

(1) التعريفات الشاملة **Inclusive** : تنظر هذه التعريفات إلى الشخصية على أنها مجموع سمات أو وحدات سلوكية متنوعة مستقلة. لذلك فإنها تميل إلى النظر إلى الشخصية نظرة جزئية، وأن الشخصية محصلة وليست مظهراً متكاملاً نسبياً ينبثق من عملية

النمو. ولعل تعريف برينس Prince يعد أحد هذه التعريفات، حيث يرى أن الشخصية هي المجموع الكلى لموروثات فطرية، ونزعات، وبواعث، وغرائز الفرد، وكذلك نزعاته المكتسبة نتيجة الخبرة .

(2) التعريفات التكاملية Integrative : تركز هذه التعريفات على الجوانب التنظيمية للعوامل التي يتكون منها سلوك الفرد. وتنظر إلى الحالات والعمليات الداخلية في ضوء العلاقات القائمة بينها. ومن هذه التعريفات تعريف وارن، وكارمايكل Warren & Carmichael، وهو أن الشخصية هي التنظيم العقلى الكامل للفرد في أى مرحلة من مراحل نموه، وهذا يتضمن : العقل، والمزاج، والمهارة، والخُلق، وجميع الاتجاهات التي تتكون خلال حياة الفرد .

(3) التعريفات الهرمية Hierarchical : تؤكد هذه التعريفات الجوانب التنظيمية للشخصية كما في التعريفات التكاملية، ولكن مدخلها في التنظيم رأسي وليس أفقياً. ومن أمثلة ذلك التعريف المبكر لماكدوجل McDougal الذي يرى أن هناك مرحلتين هرميتين للعواطف Sentiments، إحداهما بنائية، والأخرى تتعلق بتكوين عواطف أساسية في كل متناغم .

(4) التعريفات الخلْقية Idiosyncratic : تُركز هذه التعريفات على تفُرد الشخص في علاقته بالآخرين، وتحاول في معظم الأحيان التمييز بين الشخص وثقافته. إذ يرى سشون Schoen أنه إذا تصرف، وفكر، وشعر جميع أفراد مجتمع معين بنفس الطريقة، فلن يكون هناك مجال للحديث عن الشخصية. لذلك يُعرِّف الشخصية بأنها النظام المرتب، والوظيفية الكلية للعادات، والنزعات، والعواطف التي تميز الفرد عن غيره من أفراد جماعته .

(5) التعريفات التكيُفية Adaptive : تؤكد هذه التعريفات قدرة الفرد على التعامل مع بيئته بفاعلية. إذ يرى باودن Bowdn أن الشخصية هي نزعات التكيف المستقرة لدى الفرد التي تحكم سلوكه في بيئته .

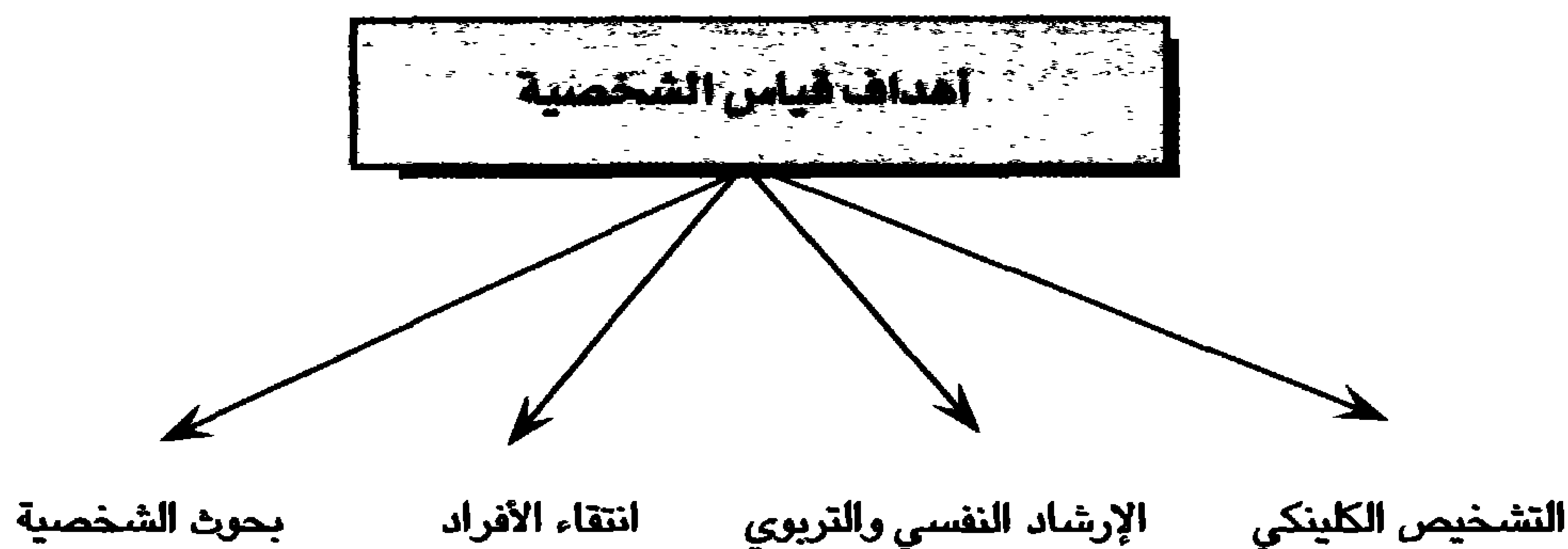
ومن هذا يتضح تباين وجهات النظر التي بينتها التعريفات الخمسة السابقة. وربما يرجع ذلك إلى تباين الأطر النظرية التي تبنَّاها علماء النفس في مجال الشخصية. ومما لا شك فيه أن هذا ينعكس على أساليب وطرق قياس الشخصية، حيث إن كل اختبار أو مقياس يستند

إلى نظرية معينة، ويُمثِّل مفاهيمها ومبادئها. لذلك يصعب الحصول على اختبار أو مقياس للشخصية لا يستند في بنائه استناداً صريحاً أو ضمنياً إلى اطار نظرى معين، حيث يُعد هذا الإطار بمثابة مرجعية الاختبار، والمحك المحدد لانتقاء المفردات أو المهام التي يشتمل عليها .

لذلك ينبغي على من يستخدم اختبارات ومقاييس الشخصية، أو يقوم بتفسير درجاتها، أن تكون لديه معرفة كافية بالنظرية التي استند اليها بناء الاختبار أو المقياس. وكذلك ينبغي أن يكون مدركاً للإطار النظري الذي تبنّاه مصمم الاختبار أو المقياس مما جعل الاختبار يقيس بعض السمات دون الأخرى .

أهداف قياس الشخصية :

كثير من المشتغلين بعلم النفس في مختلف المجالات التطبيقية، أو مجالات البحث يهتمون اهتماماً مباشراً بقياس الشخصية الإنسانية. ويوضح الشكل التخطيطي (8 - 1) التإلى هذه المجالات:



شكل (8 - 1): يوضح أهداف قياس الشخصية

التشخيص الكلينيكي :

نحتاج إلى قياس الشخصية في التشخيص الكلينيكي في مستشفيات الأمراض العقلية، والعيادات النفسية، وذلك لكي يقرر الأخصائي مدى وطبيعة الاضطراب في الشخصية الذي يعاني منه المريض، ويحدد كيفية العلاج المناسب .

الإرشاد النفسى، والعلاج النفسي :

يسهم الإرشاد والعلاج النفسى في علاج اضطرابات الشخصية بأساليب سيكولوجية. إذ يستخدم كل منها التفاعل اللفظى بين المريض والأخصائى النفسى بغرض مساعدة المريض في تغيير سلوكه لكى تتيسر له عملية التكيف. واستخدام مقاييس الشخصية قبل العلاج، وفى أثنائه، وبعد انتهائه، يصاحب عادة التشخيص الكلينيكي، والغرض من هذا القياس هو تقييم مدى التغير الذي ينتج عن أسلوب العلاج .

إنتقاء الأفراد:

تُستخدم اختبارات ومقاييس الشخصية في انتقاء الأفراد، أو ترقيتهم في وظائف أو أعمال معينة، أو قبولهم في برامج تدريبية متخصصة، مثل: تدريب الممرضات، والأطباء، ورجال البوليس، وغيرهم لزيادة فاعليتهم في أدائهم لأعمالهم .

والمشكلة الرئيسة في هذه الحالة هي ما إذا كان المتقدمون يتميزون بسمات شخصية معينة يتطلبها العمل المرشحون له، وهو ما يمكن تحديده باستخدام مقاييس الشخصية.

بحوث الشخصية:

تُستخدم بحوث الشخصية في المواقف العملية التي تتطلب مثلاً تقدير أثر برنامج تدريبى معين على الشخصية، كما تُستخدم في البحوث المختبرية، حيث تكون عينة البحث مجموعة من الأفراد المتطوعين للاشتراك في التجربة. فمثلاً يمكن إجراء تجربة عن أثر الحرمان من الطعام لفترة طويلة على شخصيات مجموعة من الأفراد .

كما تُستخدم بحوث قياس الشخصية في التحقق من كفاية طريقة القياس ذاتها. وهذا النوع من البحوث يصمم للكشف عما إذا كانت أداة قياس معينة تقيس سمات شخصية بطريقة متسقة، وما إذا كانت السمات التي تقيسها هي ما يود الباحث قياسه بالفعل. إذ ربما تدعو حاجة مؤسسة صناعية معينة، أو مستشفى معين إلى بناء مقياس جديد للشخصية لأغراض معينة، فهنا يقوم فريق البحث بدراسة مدى تحقيق المقياس للغرض الذي أعد من أجله .

وكذلك تُستخدم بغرض الإسهام في زيادة المعرفة وتطوير النظريات المتعلقة بالوظائف

النفسية، أى تطوير نظام منطقى يمكن فيه قياس التكوينات الفرضية Constructs التي يمكن الاستدلال منها على العلاقات القائمة بين الأنماط السلوكية المختلفة .

ومن هذا يتضح أن هناك هدفين رئيسين لقياس الشخصية أحدهما نظرى تطويري، والآخر عملى تطبيقى. فالهدف النظرى التطويرى يتعلق بالإسهام في التعرف على متغيرات ومفاهيم نظرية للشخصية، ومن خلال القياس يمكن ربط هذه المفاهيم بأنماط السلوك الملاحظ. أما الهدف العملى التطبيقى فإنه يتعلق باتخاذ قرارات وتنبؤات عن الأفراد في مواقف عملية.

بعض مشكلات قياس الشخصية :

يختلف قياس الشخصية عن قياس الذكاء أو غيره من الجوانب المعرفية في أن قياس الشخصية وغيرها من الجوانب غير المعرفية يكون أكثر تعقيداً. فالفرد الذي يُطبق عليه اختبار أو مقياس الشخصية لا يكون واضحاً لديه السمات أو الخصائص التي يجرى قياسها، أو كيفية تفسير الدرجات. كما أن الاستجابات تكون عرضة لتحيزات مختلفة. وكذلك ربما تؤثر عملية القياس أو الموقف الاختباري في سلوك الفرد المختبر مما قد يؤدي إلى تغيير استجابته في الاختبار، وبالتالي يجعل صدق الاختبار موضع شك.

غير أن تحيز استجابات الفرد، أو اتجاهه العقلي Response Sets، تعد من المشكلات التي لا توجد، أو على الأقل توجد بدرجة محدودة في قياس الجوانب المعرفية أو الأداء الأقصى. والمقصود بالاتجاه العقلي في الاستجابة ميل أو نزعة الفرد للاستجابة في اتجاه معين بغض النظر عن محتوى المفردة أو الفقرة الاختبارية. وكذلك نود عند قياس الانبساط مثلاً Extraversion معرفة ما إذا كان الفرد الذي يُطبق عليه المقياس يسلك بطريقة تُعبِّر عن الانبساط، وليس مجرد الموافقة أو عدم الموافقة على عبارات معينة يتضمنها هذا المقياس.

فبعض الأفراد يسلك سلوك الانبساط في مواقف وأوقات معينة دون سواها. ولذلك إذا طُلب من أحد الأفراد أن يسلك سلوكاً يدل على الانبساط، فإنه يستطيع ذلك. ولكننا لا نريد قياس هذا السلوك، وإنما نود تحديد كيف يسلك هذا الفرد عادة. لذك فإن اختبارات ومقاييس الشخصية يجب أن تُبنى بحيث تجعلنا نتأكد أن استجابات الأفراد تدل على سلوكهم النمطى أو أدائهم المميز .

وتوجد ثلاثة أنواع رئيسة من الاتجاهات العقلية هي:

(1) الإذعان أو الموافقة Acquiescence: اهتمت كثير من الأبحاث بدراسة هذا النوع من تحيز الاستجابات. ويُقصد بالإذعان أو الموافقة اتجاه الفرد أو نزعته للاستجابة بـ "نعم"، أو بـ "موافق " في اختبارات الشخصية. والدرجات التي يحصل عليها مثل هؤلاء الأفراد تكون غير صادقة، لأنه يُفترض أنها لا تقيس الاستجابات لمحتوى فقرات الاختبار، بقدر ما تقيس هذه النزعة شبه الشعورية للموافقة على فقرات أو عبارات الاختبار .

ويوجد تعريف آخر للإذعان أو الموافقة يختلط أحياناً بالتعريف السابق، وهو أنه يُعد سمة شخصية عند الأفراد الذين يستجيبون في معظم الأحيان لفقرات أو عبارات استبيانات الشخصية. ويختلف علماء النفس فيما إذا كانت هذه الاستبيانات يجب أن تقيس هذه السمة.

(2) مراعاة القبول الاجتماعي Social Desirability: كثير من الأفراد سوف يوافقون بدرجة كبيرة على عبارة مثل: «أنا شخصي ذكي ومبدع»، أما عبارة مثل: «أشعر بصداع شديد إذا سارت الأمور على غير توقعاتي»، فإنها تعد بالنسبة لهم غير مرغوبة اجتماعيا، وبالتالي سوف لا يوافقون عليها. وهذا يرجع إلى أن الفرد يحاول بطريقة شبه شعورية أن يكون مقبولا اجتماعياً. وقبوله العبارة الأولى لا يعد نوعاً من الكذب، لأن الاستجابة تكون بطريقة شبه شعورية، في حين أن الكذب يكون متعمداً.ويعتمد الفرد في استجابته على محتوى العبارات أو الفقرات، ويكون مدركاً للعلاقة بين محتوى العبارات وطريقة تصحيحها. ويمكن الاستناد إلى أحكام مجموعة من الخبراء في تقدير قيمة وزنية للقبول الاجتماعي لكل عبارة من عبارات استبيان الشخصية.

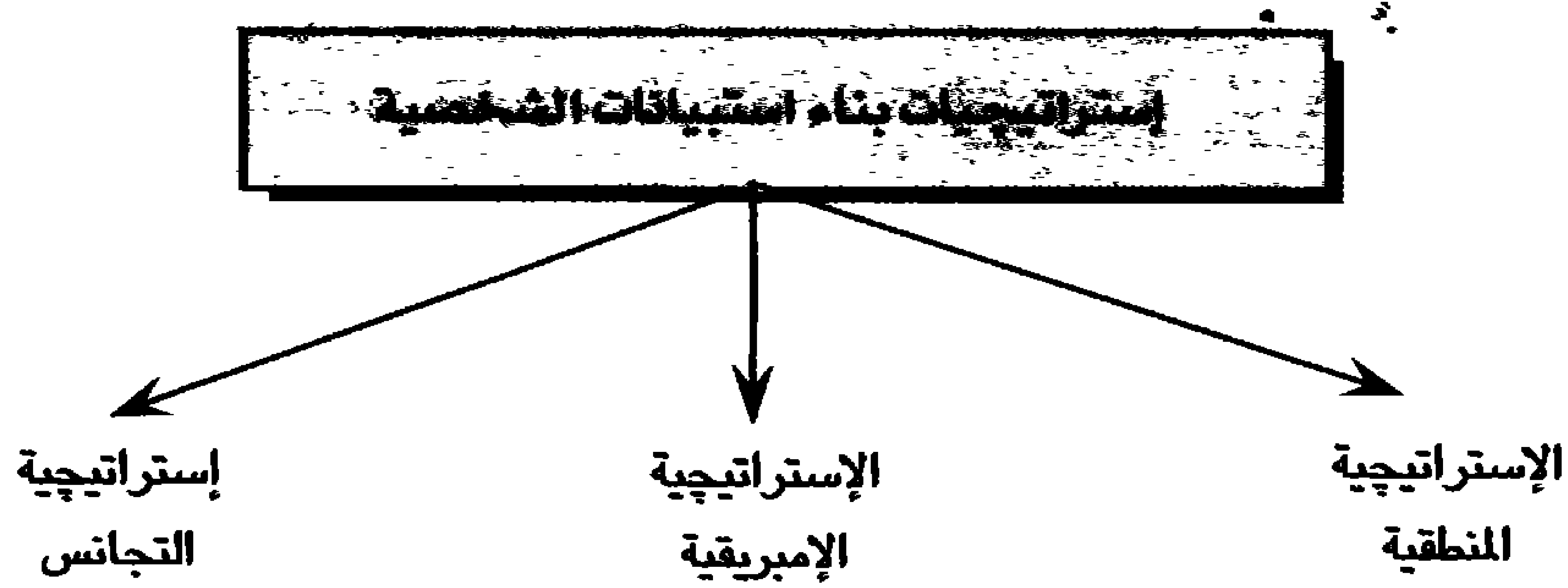
(3) الاستجابات الشاذة Deviant Responses : فهناك نزعة لدى بعض الأفراد أن يستجيبوا استجابات شاذة Atypical على فقرات أو عبارات مقاييس الشخصية، فنحن نتوقع من الأفراد غير الأسوياء أن يستجيبوا لمحتوى المقياس استجابات غير نمطية .

ويفترض بيرج Berg أن الاستجابة الشاذة للأفراد غير الأسوياء لا تعتمد على محتوى فقرات أو عبارات مقياس الشخصية. ولذلك فإنه يُفترض أنه يمكن ببساطة اكتشاف هذا الشذوذ عن طريق معرفة عدد الاستجابات الشاذة التي يختارها الفرد بغض النظر

عن محتوى الفقرات أو العبارات. ومعظم البحوث التي تهتم بدراسة الاستجابات الشاذة تتطلب من الفرد أن يختار استجابته على بُعد يتروح بين " أوافق جداً، " ولا أوافق على الإطلاق ". وهنا يلجأ الفرد غير السوى إلى اختيار استجابة تختلف عن استجابات معظم الأفراد الآخرين. ويرى بيرج أن هذه الاستجابة تتكرر في جميع المقاييس الأخرى للشخصية.

استراتيجيات بناء استبيانات الشخصية :

يعتمد كثير من مقاييس الشخصية على أساليب التقرير الذاتى Self Report التي تتطلب أن يستجيب الفرد لعبارة واضحة ومكتوبة إما (بنعم)، أو (لا). ويستند بناء هذه المقاييس إلى ثلاث استراتيجيات رئيسة يوضحها الشكل التخطيطي (8 - 2) التالي:



شكل (8 - 2): يوضح إستراتيجيات بناء استبيانات الشخصية

(1) الإستراتيجية المنطقية Logical Strategy: استُخدمت هذه الإستراتيجية في بناء المقاييس المبكرة للشخصية. فقد بُنيت على أساس نظري أو منطقي أو حدْسي. وهذا يعنى أنه بمجرد تحديد السمة أو السمات المراد قياسها تُصاغ فقرات أو عبارات تبدو أنها تقيس هذه السمة أو السمات. فمثلاً يمكن أن يشتمل المقياس الذي يقيس الانطواء على فقرات أو عبارات مثل:

"أخجل بسهولة" .

" أفضل المشاركة في الحفلات الصغيرة مع الأصدقاء على المشاركة في الحفلات الكبيرة التي تجمع عدداً كبيراً من الناس" .

"اعتبر القراءة من هواياتى المفضلة".

ويتم تصحيح الفقرات على أساس منطقي. وعلى الرغم من أنه ربما لا يكون هناك اتفاق عام على تقدير درجة فقرة معينة، إلا أنه عند بناء المقياس يجب الاهتمام بانتقاء المفردات التي يسهل تصحيحها أو تقدير درجاتها. وبناء المقياس استناداً إلى هذه الاستراتيجية يُعد خطوة أوّلية، وذلك لأن الفقرات التي تبدو أنها تقيس سمة معينة من حيث محتواها، ربما لا تقيس هذه السمة في الحقيقة، وهذا يحد من استخدامها.

(2) الاستراتيجية الإمبريقية **Empirical Strategy** : تُنتقى الفقرات أو العبارات في هذه الحالة على أساس علاقتها التجريبية بمحك خارجي. والخطوة الأولى هي اختيار مجموعة محكة Criterion Group تتميز بالسمة المراد قياسها.

فإذا افترضنا مثلاً أننا نود قياس سمة " القيادة **Leadership** "-إذا كانت القيادة هي في الحقيقة سمة نفسية = فإننا نحدد تعريف هذه السمة. ونختار مجموعة من الأفراد الذين ينطبق عليها هذا التعريف. فالقائد ربما يمكن اعتباره الشخص الذي تنتخبه الجماعة لقيادة هيئة أو جمعية معينة. وهنا يبرز التساؤل: ما الخصائص التي تميز أعضاء مجموعة القادة عن غيرهم من الأفراد؟

وإجابة هذا التساؤل تحتاج إلى تصميم عينة من المثيرات (الفقرات)،ومجموعة مقارنة تتكون من أشخاص لا يتميزون بسمة القيادة.

وربما يتطلب تصميم وبناء عينة الفقرات الاسترشاد بالدراسات المتعلقة بالقيادة، ومراجع علم النفس الاجتماعي، وغير ذلك. كما يمكن أن نطلب من مجموعة من الخبراء كتابة قائمة بالصفات التي يرون أنها تميز القائد، وكذلك الاطلاع على المقاييس الأخرى التي تقيس القيادة والسمات المتعلقة بها .

وبعد كتابة عينة ممثلة من الفقرات، نقوم بتطبيقها على مجموعة القادة، ومجموعة أخرى لا تتصف بالقيادة. ولكي نحدد أي الفقرات تميز بين المجموعتين، يمكن إيجاد نسبة عدد الافراد الذين استجابوا لكل مفردة بطريقة معينة في كل من المجموعتين، ومن ثم نختار الفقرات التي تميز بدرجة كبيرة بينهما .

ومن الجدير بالذكر أن مقياس الشخصية الذي يُبنى استناداً إلى هذه الاستراتيجية يمكن أن يشتمل على فقرات أو عبارات ربما لا يبدو أن محتواها له علاقة منطقية بالسمة أو السمات المراد قياسها .

غير إنه ينبغي مراعاة أن انتقاء الفقرات على أساس نتائج مجموعة محكة معينة ربما تعكس طبيعة وخصائص هذه المجموعة على المقياس. فاختيار مجموعة القادة في هذا المثال وفقاً لتعريف معين ربما لا يكون كافياً. وإذا اختلف التعريف، أو لم تحقق المجموعة المحكة صفات القيادة المذكورة في التعريف الذي نختاره، فإن المقياس لا يكون مناسباً. كما ينبغي مراعاة أن الاستناد إلى هذه الاستراتيجية في بناء مقاييس الشخصية ربما يضمن أن تتميز هذه المقاييس بالصدق التلازمي، ولكن يبرز التساؤل حول مدى الاستفادة من المقياس في التنبؤ بمحك يختلف عن مجرد قياس انتماء الفرد لعضوية المجموعة المحكة الأصلية ؛ فهذا التساؤل يتعلق بالأهمية النظرية للمقياس.

استراتيجية التجانس **Homogeneity Strategy**

يرى بعض علماء القياس أن المقاييس التي تتميز بتجانس مفرداتها هي التي يمكن استخدامها في قياس متغيرات سيكلوجية ذات معنى. أى أن الأساس في المقياس أن يكون متجانساً، وهذا يعني أن الفقرات التي لا ترتبط بغيرها من فقرات المقياس يجب حذفها، لأنه يمكن اعتبارها في هذه الحالة أنها تقيس سمة أو تكويناً فرضياً مختلفاً .

ويتطلب بناء مقياس الشخصية استناداً إلى هذه الاستراتيجية البدء بعدد كبير من الفقرات، وتطبيقها على عينة تقنين مناسبة من الأفراد، وتُحسب معاملات الارتباط بين المفردات، ويُستخدم أسلوب التحليل العاملي Factor Analysis – وهو من الأساليب الإحصائية متعددة المتغيرات سبقت الإشارة إليه في الفصل الرابع – لتجميع الفقرات في أقل عدد من العوامل المتعامدة التي تفسر التباين في مصفوفة معاملات الارتباط.

وبذلك نكون قد صنَّفنا الفقرات في مجموعات متجانسة، تقيس كل مجموعة منها سمة معينة يشتمل عليها المقياس. ويُعطى اسماً للمقياس في ضوء محتوى الفقرات التي تشتمل عليها كل مجموعة .

ومقاييس الشخصية التي تستند في بنائها إلى هذه الاستراتيجية تكون أحادية البُعد Uni-dimensional، وتتميز بقدر من صدق التكوين الفرضي Construct Validity الذي أوضحناه في الفصل الرابع .

ومن الجدير بالذكر أن هذه الاستراتيجيات الثلاث في بناء مقاييس الشخصية التي تعتمد على التقرير الذاتي باستخدام الاستبيانات، تعد مداخل رئيسة في هذا الشأن. وعلى الرغم من أن الطريقة المنطقية أقلها استخداماً، إلا أنه يمكن الاسترشاد بها في انتقاء المفردات في المراحل الأولية لبناء المقياس، لكي يكون للمفردات أساس منطقي، وربما يُفضل الجمع بين الاستراتيجية الإمبريقية، واستراتيجية التجانس في بناء مقاييس الشخصية، حيث نحاول بناء مقياس فقراته متجانسة، ثم نُقَدِّر صدقه عن طريق محك تجريبي، ونستبعد الفقرات التي لا تحقق هذا المحك. وبذلك يصبح للمقياس أهمية نظرية وفائدة عملية.

أنواع اختبارات ومقاييس الشخصية:

يمكن قياس الشخصية بأنواع مختلفة من الاختبارات والمقاييس؛ فقياس وتقييم الشخصية يتضمن الملاحظة الموضوعية المنظمة لسلوك الفرد تحت شروط محددة في علاقتها بمثيرات معينة. وللحصول على بيانات تتعلق بفرد ما، يجب محاولة فهم أثر مختلف مكونات الموقف المثير على سلوكه .

ويتحدد الموقف المثير بخصائصه الفيزيائية، وهو ما يجد الفرد نفسه في مواجهته، وبالتعليمات التي تُعطي لفرد، وبالمطلب المحدد (المثيرات) التي يواجهها. فنحن نستطيع أن نُكوِّن فكرة واضحة عن طبيعة الشخصية، ونجمع بيانات تستند إلى نظرية معينة من خلال ملاحظة *استجابات الأفراد لمثل هذه المواقف المحددة تحديداً واضحاً* .

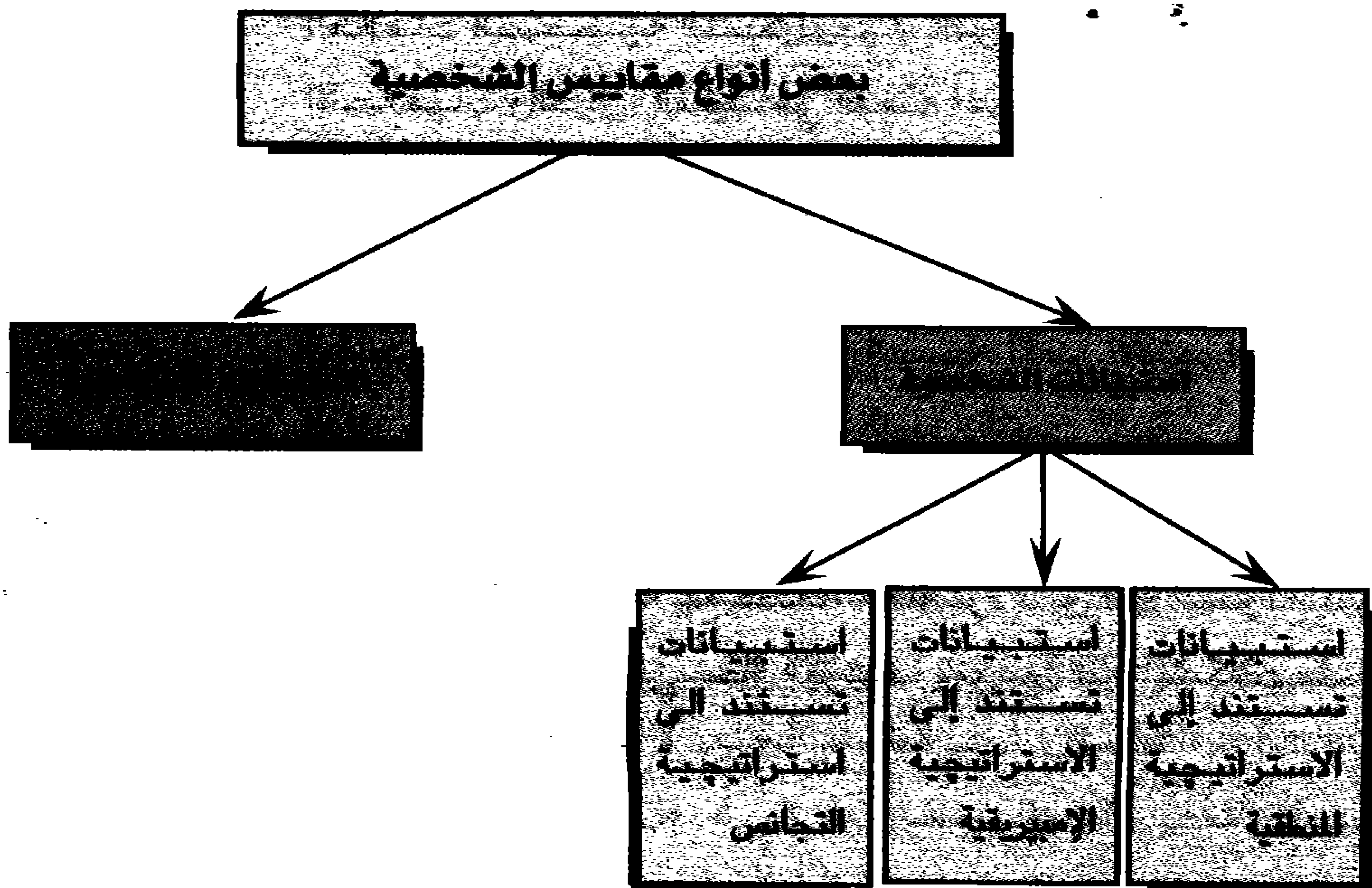
ونظراً لاختلاف المتغيرات المرتبطة بعملية جمع البيانات المتعلقة بشخصية الفرد، فإن هناك تبايناً واضحاً في طرق وأساليب جمع هذه البيانات، من حيث: طبيعة الموقف الاختباري، وطبيعة المثيرات، وطبيعة التعليمات المعطاة، وطبيعة الاستجابة المطلوبة، وكيفية تقدير الدرجات، وتحليلها، وتفسيرها .

واختلاف نظريات الشخصية يؤدى إلى اختلاف تعريفات مفهوم الشخصية كما سبق أن أوضحنا، وبالتالى يؤدي إلى اختلاف طرق وأساليب القياس والتقدير. كما تؤدي إلى طرق

مختلفة لملاحظة سلوك الأفراد. وهذا يتطلب بالضرورة فهم العلاقات القائمة بين نظريات الشخصية وأساليب القياس والتقدير. فنظرية الشخصية التي لا تسهم في تحديد وتيسير وسائل جمع معلومات عن الأفراد تُعد نظرية غير ذات جدوى من الناحية العملية. كما أن البيانات التي لا تتعلق بنظرية معينة تكون عديمة المعنى، وقليلة الفائدة.

وينبغي أن تكون الطريقة المستخدمة في قياس الشخصية أكثر حساسية للسمة المراد قياسها، ويجب ألا تتأثر بأي متغيرات أخرى لا ترتبط بهذه السمة. وهذا يتطلب تحديد الهدف من القياس، وماذا نقيس، وفيم تستخدم النتائج.

ويمكن تصنيف مقاييس الشخصية التي سأتناولها في هذا الفصل في ثلاثة اقسام رئيسة وفقا للاستراتيجية المستخدمة في بنائها، كما هو موضح بالشكل التخطيطي (8-3) التالي:



شكل (8-3): يوضح أنواع مقاييس الشخصية

وسوف أقدم فيما يلي بعض الأمثلة لكل نوع من أنواع اختبارات ومقاييس الشخصية الموضحة بهذا الشكل:

أولاً: استبيانات تستند إلى الاستراتيجية المنطقية:

قائمة مراجعة المشكلات لموني Mooney Problem Check- List:

تتضمن هذه القائمة أربعة مستويات:

المستوى الأول: للصفوف من السابع إلى التاسع، وينتج عن القائمة درجات تتعلق بالصحة والنمو الجسمي، المدرسة، المنزل والعائلة، المال، العمل، المستقبل، العلاقات الاجتماعية، أمور تتعلق بذات الفرد.

المستوى الثاني: للصفوف من التاسع إلى نهاية المرحلة الثانوية، ويتناول مشكلات تتعلق بالصحة والنمو الجسمي، النواحي المالية، ظروف المعيشة، التوظيف، الأنشطة الاجتماعية والترويحية، العلاقات الاجتماعية السيكلوجية، العلاقات الشخصية السيكولوجية، التودد، الزواج، المنزل والعائلة، الأخلاقيات والتدين، التكيف في البيئة المدرسية، المستقبل الدراسي والمهني.

المستوى الثالث: للمرحلة الجامعية، ويتناول نفس مشكلات المستوى الثاني .

المستوى الرابع: للراشدين، ويتناول مشكلات تتعلق بالصحة، الأمن الاقتصادي، التطوير الذاتي، الشخصية، المنزل والأسرة، التودد، التدين، المهنة.

وقد اعتمد بناء فقرات هذه القائمة على تحليل دقيق لعدد كبير من العبارات المتعلقة بمشكلات الطلاب والراشدين صاغتها عينة بلغ عددها 4000 فرد تمثل هذه الشرائح الأربع، وكذلك على الأدبيات، وخبرات معد القائمة في مجال الإرشاد. ومن بين هذه المشكلات ما يلي:

- أود أن أتعلم كيف أقرأ بطريقة أفضل.
- لا أعيش مع والديً.
- أعاني من مشكلات في أسناني.

ويمكن في كل مستوى من المستويات الأربعة أن يوضح الفرد ما إذا كان يود مناقشة مشكلاته مع شخص معين يحدده بنفسه. وهذه القائمة ليست اختبارا بالمفهوم السيكومتري، حيث إن الفرد يستجيب للعبارة التي يشعر أنها تُعبر عن مشكلة لديه. لذلك فإن القائمة تفتقر

إلى بيانات تتعلق بالثبات والصدق، على الرغم من أن دليل القائمة أشار إلى بعض الدراسات التي أجريت في هذا الشأن. والهدف من هذه القائمة هو تقديم معلومات مساعدة أو إضافية للطالب أو الراشد، وليس لعباراتها إجابات صحيحة أو خطأ، وإنما تحدد مشكلات ، أو مجالات اهتمام فقط.

ومن الجدير بالذكر أن قائمة التفضيل الشخصي التي أعدها إدواردز Edwards Personal Preference Schedule (EPPS) تعد أحد استبيانات الشخصية شائعة الاستخدام، واستندت في بنائها إلى الاستراتيجية المنطقية أيضاً، ولا تعد اختباراً شأنها شأن قائمة موني، لأن عباراتها ليس لها إجابة صحيحة أو خطأ.

وتستند القائمة إلى الإطار النظري لنظام الحاجات عند موراي Murray الذي أثر تأثيراً كبيراً في بناء اختبارات الشخصية.

ثانيا: استبيانات تستند إلى الاستراتيجية الامبريقية:

استبيان مينسوتا للشخصية متعدد الاوجه
Minnesota Multiphasic Personality Inventory (MMPI):

يعد هذا الاستبيان من أكثر استبيانات الشخصية التي تستند إلى التقرير الذاتي، والاستراتيجية الامبريقية، شيوعا واستخداماً، حيث أجريت حوله عدة آلاف من الدراسات والبحوث. وقد أعد هذا الاستبيان كل من هاثاواي Hathaway، وماكنليMckinley بجامعة مينسوتا الامريكية عام 1942 بهدف التشخيص الكلينيكي لمرضى العيادات النفسية. ويشتمل الاستبيان على (556) عبارة من نوع الصواب أو الخطأ، وتشير جميعها إلى الفرد Self-referenced مثل: أحب الطعام الجيد، لا أعاني صعوبة عندما أريد النوم، أصرخ بسهولة، أشعر بالخوف عندما أنظر إلى أسفل من مكان مرتفع. وقد تم تصنيف عبارات الاستبيان في عشرة مقاييس فرعية كلينيكية، وأربعة مقاييس لضبط مدى صدق إجابات الفرد على العبارات.

وهذه المقاييس الكلينيكية كالتالي:

- توهم المرض Hypochendriasis .
- الانقباض Depression.

- الانحراف السيكوباتي Psychopathic Deviate.
- الذكورة – الأنوثة Masculinity - Feminity.
- البارانويا Paranoia.
- الوهن النفسي Psychasthenia.
- الفصام Schizophrenia.
- الهوس الخفيف Hypomania .
- الانطواء الاجتماعي Social Introversion.

ولبناء الاستبيان بدأ العالمان باختيار حوالي (2000) عبارة من بطاقات الفحوص الكلينيكية، ومراجع الطب النفسي، وأساليب الفحص العصبي، ومن مقاييس الاتجاهات الشخصية والاجتماعية التي سبق بناؤها. ثم قاما بتطبيق هذه العبارات على حوالي (200) من المرضى المصابين بأمراض عصبية نفسية في مستشفيات جامعة مينسوتا. كما طبقت على حوالي (1000) من الأفراد الأسوياء الذين تم اختيارهم من بين الذين كانوا يزورون أقاربهم المرضى في هذه المستشفيات. وطبقت العبارات أيضاً على عدة مئات من الطلاب المتقدمين للالتحاق بجامعة مينسوتا، وكذلك على عينة من سكان إحدى المدن التابعة لهذه الولاية الأمريكية. وقد تم اختيار العبارات التي اشتمل عليها الاستبيان في صورته النهائية على أساس تمييزها بين هؤلاء المرضى والأسوياء.

وتغطى هذه العبارات مدى واسعاً من الموضوعات التي تتناول جوانب الشخصية المختلفة، مثل: الجوانب الصحية، والاتجاهات الجنسية والاجتماعية، والدينية، والسياسية، والأسرية، والمهنية، والمخاوف المرضية، والحالات الانفعالية المختلفة، والروح المعنوية، وما يتصل بالذكورة والأنوثة، واتجاه الفرد نحو الاختبار.

ويُستخدم هذا الاستبيان في صيغته العربية في التشخيص الكلينيكي لمن هم في مستوى المرحلة الثانوية وما بعدها، ويمكن استخدامه كاختبار فردي، أو جماعي.

ولتفسير الدرجات الخام في جميع المقاييس الفرعية، ومقاييس ضبط الصدق ينبغي تحويلها إلى درجات تائية متوسطها (50)، وانحرافها المعياري (10) .

وقد وُجِّه بعض النقد إلى هذا الاستبيان لانخفاض ثبات بعض المقاييس التي يتكون منها. وقد برَّر المؤلفان ذلك بأن الطرق السيكومترية المتعارف عليها في تقدير الثبات لا تصلح لاختبارات الشخصية، وحجتهما في ذلك أن سمات الشخصية سريعة التغير، وأن البيانات المستمدة من إعادة تطبيق الاستبيان تعد مقياساً لتباين السمات، وليست تقديراً لثبات الدرجات، حيث يتوقع حدوث هذا التباين في السمات لدى من يعانون من أمراض نفسية.

ومن المقاييس الأخرى المعروفة التي استندت في بنائها إلى الاستراتيجية الامبريقية استبيان كاليفورنيا السيكلوجي California Psychological Inventory (CPI).

ويشتمل هذا الاستبيان على (480) عبارة يشبه أكثر من ثلثها عبارات استبيان مينسوتا متعدد الأوجه، وجميعها من نوع الصواب أوالخطأ. ومن ميزات هذا الاستبيان أنه يطبق على الأفراد العاديين، وأجريت حوله كثير من الدراسات التطويرية لإثراء استخداماته، وزيادة صدقه المتعلق بالتكوين الفرضي.

(ثالثاً) استبيانات تستند إلى استراتيجية التجانس:

استبيان عوامل الشخصية الستة عشر لكاتل :Cattell 16 - PF Questionanire:

استخدم كاتل Cattell أسلوب التحليل العاملي في بناء هذا الاستبيان. وقد سبق أن ذكرنا أن هذا الاسلوب يعد طريقة إحصائية لإرجاع العدد الكبير من معاملات الارتباط الملاحظة بين المتغيرات أو الاستجابات المختلفة لفقرات أو عبارات الاختبارات، أو الاستبيانات، إلى عدد قليل من العوامل غير المترابطة. والفقرات التي تضم في تجمع معين ترتبط ببعضها ارتباطاً عالياً، وترتبط بغيرها من الفقرات التي في التجمعات الأخرى ارتباطاً منخفضاً، أو لا ترتبط بها على الاطلاق. وبعد تحديد الفقرات التي تُكَوِّن التجمعات أو العوامل المختلفة، يمكن استخدامها كميزان Scale لقياس سمات الشخصية. ويقوم معد الاختبار أو الاستبيان بإعطاء اسم للميزان في ضوء محتوى الخاصة أو السمة التي يتضح أن الفقرات تشترك فيها. ونظراً لأن هذه التسمية تعتمد على الأحكام الشخصية، فإنه من المحتمل أن يصل باحثون مختلفون إلى تسميات مختلفة للميزان نفسه. غير أن الفقرات وليست التسمية في ذاتها هي التي تُستخدم كتعريف للميزان.

وقد أجرى كاتل تحليلاً عاملياً لاستبيان يشتمل على (80) متغيراً طبقه على عينة تشتمل على (370) فرداً، وتمكن من استخلاص (20) عاملاً. ولكنه استبعد عدداً من هذه العوامل لعدم وضوحها عندما أعاد النظرفي بناء المقياس، وتبقى لديه (15) عاملاً من عوامل الشخصية، وعامل يدل على الذكاء العام. وقد أضاف كاتل بعض الفقرات الأخرى على أساس ارتباطها بهذه العوامل الستة عشر. وبذلك أصبح عدد فقرات الاستبيان (186) فقرة. ويتكون من صيغتين متكافئتين، وكل عامل منها يقاس بعدد يتراوح بين (10 ، 13) فقرة.

وأجرى كاتل مرة أخرى تحليلاً عاملياً على الاستبيان بعد إضافة عدد آخر من الفقرات لكل من صيغتيه المتكافئتين. ومن أمثلة بعض عباراته: "أحب مشاهدة الألعاب الرياضية الجماعية"، ويستجيب الفرد للعبارات إما بكلمة "نعم" أو "أحياناً"، أو "لا". ويُنصح الفرد بالاستجابة بأول انطباع يأتي لديه، بحيث تكون استجابته للموقف الذي تصفه العبارة هي استجابته لمعظم المواقف التي تشبه ذلك الموقف. كما يُنصح بعدم الاستجابة بكلمة "أحياناً" إلا إذا لم يستطع الاختيار بين البديلين الآخرين.

ويمكن اعتبار هذا الاستبيان من نوع المقابلة المقننة المنتظمة غير الشخصية. وهو محدد البنية، لأن كل فقرة أو عبارة لها ثلاث استجابات ذاتية أو حرة Voluntary Responses. كما أنه مباشر وواضح الهدف، بمعنى أن المختبر يعلم أن الاستبيان هو مقياس للشخصية. ويمكن تقدير درجاته بطريقة موضوعية يدوياً، أو باستخدام الحاسوب.

ويحصل الفرد على درجة لكل عامل من العوامل الستة عشر، حيث يفترض أن هذه تعطي صورة عن شخصيته بوجه عام. ويمكن الرجوع إلى دليل الاختبار Test Manual لمعرفة المعنى السيكلوجي لكل من هذه العوامل حيث تُفسر على أنها سمات الشخصية. ومن بين هذه السمات: التحفظ في مقابل الانطلاق، الخضوع في مقابل تأكيد الذات، الخجل في مقابل الجسارة والإقدام. ويمكن رسم صفحة نفسية (بروفيل) لدرجات الفرد تساعد في التفسير والتشخيص والتنبؤ.

وقد بينت بعض الدراسات أن ثبات درجات عوامل أي من صيغتي هذا الاستبيان منخفض بعامة بسبب قلة عدد فقرات كل من الصيغتين. كما تراوحت معاملات ثبات المقاييس الفرعية بطريقة التجزئة النصفية بين (0.93، 0.54) بينما معامل الثبات الناتج من إعادة تطبيق الاستبيان بعد حوالي أسبوع كان أقل من (0.80) أما البيانات الإحصائية المتعلقة بصدق

المقياس فلم تكن كافية للاستفادة منها في معرفة ما إذا كانت درجات الاستبيان تتنبأ بالفعل بالسلوك المطلوب قياسه، مما يستدعي مزيداً من الدراسات والبحوث للتوصل إلى أفضل طرق استخدام هذا الاستبيان في المجالات الصناعية والتعليمية والكلينيكية.

تعقيب عام على استبيانات الشخصية:

يتضح من عرضنا لاستبيانات الشخصية أن هناك من الأدلة الإمبريقية ما يبين أن اتساق درجات هذه الاستبيانات، سواء معاملات الاستقرار، أو التكافؤ، أو الاتساق الداخلي، أقل بوجه عام من غيرها من مقاييس الجوانب العقلية المعرفية. ولكن هل هذا الانخفاض يرجع إلى طبيعة الظاهرة المقاسة، أم إلى عملية القياس ذاتها؟ إن هناك من الأدلة أيضاً ما يؤكد أن الاستبيان الذي يُعتنى ببنائه استناداً إلى أي من الاستراتيجيات الثلاث السابقة يمكن أن يحقق قدراً معقولاً من الاتساق الداخلي والاستقرار في درجاته.

أما فيما يتعلق بالصدق، فإنه يبدو أن صدق التكوين الفرضي أكثر ملاءمة في معظم الأحيان لتقدير صدق استبيانات الشخصية. إلا أنه لا يمكن التوصل إلى هذا الصدق في وقت قصير، كما لا يمكن التعبير عن هذا النوع من الصدق بمؤشر واحد.

والسؤال الآخر هو: هل أمكن التوصل إلى ما يرتبط تجريبياً بدرجات استبيانات الشخصية؟ والحقيقة ان معظم البيانات التي جمعتها الدراسات المتعددة في هذا الشأن ارتبط كل منها بموقف معين، واستُخلصت من عينات مختلفة في أوقات مختلفة ولأغراض مختلفة، بدلاً من كونها جزءاً من محاولات منظمة ومتكاملة لفهم معنى الدرجات وتفسيرها. لذلك يصبح من الصعب دراستها واستيعاب دلالاتها بشكل متكامل.

أما المشكلة المهمة الأخرى فتتعلق بتزوير الاستجابات، أو ما يعرف بتحيز الاستجابات Response Bias ، والذي يبدو تأثيرها واضحاً في استبيانات الشخصية أكثر من غيرها من المقاييس. فعلى الرغم من الأبحاث المستفيضة التي اهتمت بضبط تحيز الاستجابات وبخاصة الاتجاهات العقلية في الاستجابات، إلا أنها لم تستطع استبعاد هذه المتغيرات من أي استبيان للشخصية. وعلى الرغم من أن تحديد اتجاه ومقدار تحيز الاستجابات في بعض الاستبيانات - كما في مقاييس ضبط مدى صدق الاستجابات - واستخدام هذه البيانات في عملية تفسير الدرجات يعد خطوة مطلوبة نحو التغلب على هذه المشكلة، إلا أن الحل الأفضل ربما يكون عن



طريق التوصل إلى نمط معين من أنماط الاختبار بحيث يسهم في استبعاد أثر تحيز الاستجابات.

والآن ما أكثر الاستراتيجيات فاعلية في بناء استبيانات التقرير الذاتي للشخصية؟ واجابة هذا التساؤل تتطلب معرفة الغرض الذي يُبنى من أجله الاستبيان ليكون أكثر فاعلية. فالاستبيانات التي يستند بناؤها إلى الاستراتيجية الإمبريقية (مثل استبيان الشخصية متعدد الأوجه) تكون أكثر استخداماً، وتهتم كثير من البحوث السيكولوجية بدراستها. وهذه النتيجة يمكن أن تنطبق على الاستبيانات التي تستند إلى استراتيجية التجانس (مثل استبيان كاتل للعوامل الستة عشر)، حيث يجب أن تتميز هذه الاستبيانات بمزايا الاستبيانات التي تستند إلى الاستراتيجية الإمبيريقية بالإضافة إلى أنها يمكن أن تمدنا بدرجات يسهل تفسيرها.

وبالطبع فإن أداة القياس المثالية يجب أن تشتمل على فقرات أو عبارات يتم انتقاؤها باستخدام الاستراتيجيتين معاً.

المقاييس الإسقاطية للشخصية:

إن بناء استبيانات الشخصية التي تعتمد على التقرير الذاتي لا يختلف كثيراً عن بناء مقاييس الجوانب المعرفية، مثل: التحصيل، والذكاء، والاستعدادات الخاصة، حيث تهتم أيضاً بالموضوعية والتقنين. وربما يكون الاختلاف فقط في محتوى المقياس ومحكات تقدير صدق درجاته. غير أنه وجه كثير من النقد لهذه الاستبيانات الموضوعية فيما يتعلق بتركيزها على كيفية قياس السمات الشخصية أكثر من تركيزها على ما الذي تقيسه هذه الاستبيانات. لذلك يرى بعض علماء النفس أنه ينبغي استناد مقاييس الشخصية بدرجة أكبر على النظريات السيكلوجية، ولعل المقاييس الإسقاطية للشخصية حاولت ذلك من أجل التوصل إلى بيانات متكاملة يمكن أن تعطى صورة ديناميكية لشخصية الفرد بدلاً من التركيز على أبعاد مستقلة للشخصية كما في استبيانات الشخصية. لذلك تعتمد المقاييس الإسقاطية على منهجية مختلفة، وتمثل توجهاً مغايراً لاستبيانات الشخصية.

ويمكن أن تتضح العلاقة بين الطرق الإسقاطية Projective Techniques والنظريات السيكوديناميكية Psychodynamics في السس التي بنيت على أساسها المقاييس الإسقاطية للشخصية. فالنظريات السيكوديناميكية تؤكد الجوانب المزاجية الذاتية للفرد، وتعتبر أن

السلوك يحدث نتيجة تفاعل قوي متعددة، مثل: الدوافع، والحوافز، والحاجات، والصراعات، وان تكوين الشخصية يتضمن تنظيمات مختلفة، مثل الشعور واللاشعور، وأن الشخصية هي العملية التي يفرض فيها الفرد تنظيماً وتكويناً على المثيرات الخارجية في بيئته المحيطة به.

ويمكن أن نرى العلاقة بين الطرق الإسقاطية في تقييم الشخصية، والنظريات السيكودينامكية في الشخصية في خصائص هذه الطرق التي تتمثل في:

(أ) تأكيد إعطاء حرية كافية للفرد لاختيار استجابته من بين عدد لا نهائي من البدائل، وذلك لأن الاختبار يكون غير محدد البنية وذاتي.

(ب) تعليمات الاجابة، والمثيرات التي يشتمل عليها الاختبار تعطى مؤشرات قليلة للاستجابة.

(جـ) الغرض من الاختبار يكون غير واضح بالنسبة للفرد المختبر.

(د) يهدف الأخصائي الكلينيكي الذي يستخدم هذه الطرق للقيام بتفسيرات كلينيكية وشاملة من البيانات التي يجمعها.

وعندما يستخدم المصطلح "إسقاط Projection في الطرق الإسقاطية للشخصية، فإنه يعني بعامة العملية التي تؤثر بها خصائص الفرد، أي بنية شخصيته، في كيفية إدراكه، وتنظيمه، وتفسيره لبيئته وخبراته. وهذه التأثيرات يمكن ملاحظتها بدرجة أفضل وقياسها عندما يستجيب الفرد لمثيرات أو مواقف تتميز بالغموض أو الحداثة.

وقد استخدم هنري موراي Murray عام 1938 مصطلح الإسقاط Projection لأول مرة كطريقة لتقدير الشخصية، وصمم اختباراً لهذا الغرض سنوضحه بعد قليل. ولكن يرجع الفضل في تأكيد أهمية الطرق الإسقاطية إلى فرانك Frank عام 1939؛ فقد عارض استخدام الاختبارات الموضوعية للشخصية بحجة أنها تُستخدم في تصنيف الأفراد، ولا تعطينا معلومات مفيدة عن الفرد كإنسان متمايز، وحث على استخدام اختبارات تساعد في فهم أكثر وضوحاً للمشاعر والأفكار الخاصة عند الأفراد ومثل هذه الاختبارات تسمح للفرد أن يفرض تكويناً وتنظيماً خاصاً به على المثيرات، وبهذا تُعبر عن تصور ديناميكي للشخصية، ويمكن من استجاباته التوصل إلى استدلالات فيما يتعلق ببنية شخصيته.

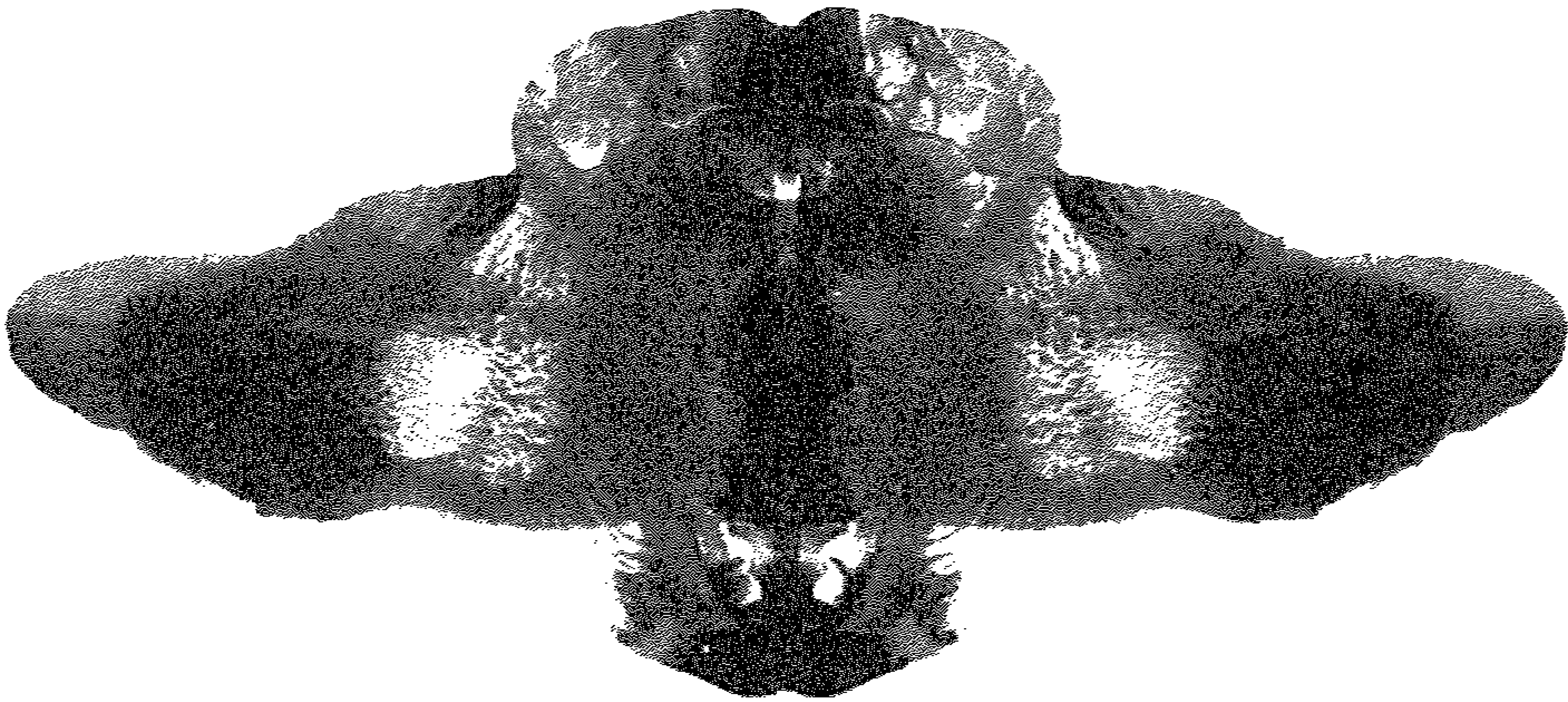
غير أنه لا يوجد اتفاق كامل بين علماء النفس حول المظاهر الرئيسة للطرق الإسقاطية. فقد

قام لينزي (Lindzey, 1961) مثلاً بإجراء مسح للتعريفات المختلفة للطرق الإسقاطية، وتوصل إلى أن الخصائص المميزة لهذه الطرق هي: حساسيتها للعمليات اللاشعورية أو المظاهر الكامنة للشخصية، وتعدد استجابات الأفراد وذلك لتعدد أبعاد الشخصية التي تقاس في آن واحد، وعدم معرفة الفرد بالغرض من الاختبار، وثراء الاستجابات التي تنتج عن المثيرات، وعدم وجود استجابة صحيحة وأخرى خطأ.

وعلى الرغم من تعدد أنواع الاختبارات الإسقاطية، الا أنني سأقتصر على تقديم مثالين فقط لاختبارين يعدان أكثر الاختبارات الإسقاطية شيوعاً واستخداماً في الأغراض الكلينيكية، والتي أجريت حولها عدة آلاف من الدراسات والبحوث أشار إلى بعض منها الكتاب السنوي الأمريكي للقياس العقلي Buros' Mental Measurment Yearbook.

(1) اختبار بقع الحبر لرورشاخ :Rorschach Inkblot Test:

صمّم هيرمان رورشاخ – الطبيب النفسي السويسري – اختبار بقع الحبر الذي سُمّي باسمه. وقد صمم مثيرات عن طريق وضع بقع الحبر على ورقة وقام بثني الورقة بحيث ينتج عنها أشكالاً متماثلة ولكن غير محددة البنية. ثم قام بتطبيق هذه المثيرات على مرضى المستشفيات العقلية، واحتفظ بالبقع التي استدعت استجابات مختلفة من مجموعات مختلفة من المرضى، واستبعد البقع الأخرى. ويوضح الشكل التالي (8-3) صورة مماثلة لإحدى هذه البقع:



شكل (8-3): يوضح صورة مماثلة لإحدى بقع الحبر في اختبار رورشاخ

وقد استخدم رورشاح في محاولاته آلافاً من بقع الحبر حتى استقر في النهاية على عشر بقع في بطاقات، وصمم طريقة لتقدير الدرجات. وتم نشر الاختبار، وطريقة تقدير درجاته، ووصف طريقة بنائه عام 1921. ولا تزال هذه البقع تُستخدم حتى وقتنا الحاضر في صورة بطاقات على كل منها بقعة حبر تتخذ شكلاً معيناً غير محدد البنية. وأعيد النظر في تقدير درجات الاختبار، ولكن الجهود التي بذلها رورشاخ في البداية ظلت أساسية في معظم نظم تقدير درجات الاختبارات المستخدمة حالياً.

ولقد كان رورشاخ متأثراً بنظرية فرويد، وما تتضمنه من مفاهيم، مثل: اللاشعور، وديناميكية الشخصية. ولكنه لم يصمم الاختبار على أساس هذه النظرية، ومع هذا فإن طبيعة بناء الاختبار يبدو أنها قد تأثرت بها بلا شك.

ويحاول الفاجص عند تطبيق بطاقات بقع الحبر أن يجعل المفحوص في حالة عدم توتر، وراحة تامة، وتكون التعليمات كافية فقط بالقدر الذي يمكنه من استكمال استجاباته. ولهذا فإنه يُطلب من المفحوص أن ينظر إلى كل بطاقة، ويذكر للفاحص ما يراه فيها.

وأهم ما تتميز به هذه الطريقة هو حرية الفرد في اختيار ما يراه، وأين يراه، ومحددات إدراكاته. ويقوم الفاحص اثناء التطبيق بتسجيل الاستجابات التي يقدمها المفحوص، كما يسجل المواضع المختلفة في البطاقة التي ركز عليها المفحوص في استجاباته، وما الذي جعل البقعة تأخذ هذا الشكل. وبذلك يمكن للفاحص ان يحدد الأساس الذي بني عليه المفحوص استجابته بدقة اكبر.

ويتركز الاهتمام عند تفسير الاستجابات بطريقة تكوين هذه الاستجابات أو المدركات Per-cepts ومحددات الاستجابة، ومحتواها. وفي جميع الحالات يُفترض أن هناك تناظراً بين طريقة الفرد في تكوين مدركاته، والطريقة التي يتعامل بها بعامة مع المثيرات في البيئة المحيطة به، ويُقصد بذلك كيف ينظم ويُشكل هذه المثيرات. فالمدركات الجيدة التنظيم، والتي تتميز بخصائص تكوينية جيدة في علاقتها بتكوين بقعة الحبر، ربما تؤدي إلى اقتراح أن الوظائف النفسية موجهة توجيهاً واقعياً، وتعمل في مستوى جيد. أما المدركات التي تكوينها غير جيد، ولا تناسب تكوين بقعة الحبر، ربما تؤدي إلى اقتراح أن المفحوص يسلك سلوكاً وهمياً، أو غير واقعي.

ويختلف تفسير شخصية الفرد تبعاً للمحتوى الذي يراه في معظم البطاقات، مثل: الحيوانات أو الانسان، أو المحتوى الذي يُعبر عن عدوان أو عاطفة. فمثلاً يمكن المقارنة بين تفسيرين ربما يصل إليهما الفاحص من مجموعتين من الاستجابات، إحداهما رأى فيها الشخص حيوانات تتصارع، والأخرى رأى فيها أناساً يشاركون مشاركة كاملة في عمل معين.

وفي كثير من الحالات يكون محتوى الاستجابات للبقع على شكل قصص، ويفسرها الفاحص على أنها تعبِّر عن خصائص شخصية المفحوص، وأسلوب حياته.

كما يمكن تفسير المحتوى بطريقة رمزية، فمثلاً: الانفجار ربما يرمز إلى عدوان شديد، والثعلب ربما يرمز إلى الميل إلى العدوان، والخنزير إلى الشره والنهم، والعنكبوت والأخطبوط والساحر والغول ربما يرمز كل منها إلى اتجاهات سلبية نحو الأب المسيطر، والنعامة ربما ترمز إلى محاولة الابتعاد عن الصراعات.

والأقسام العامة للمحتوى، مثل: الطعام، والحيوانات التي يؤكل لحمها، والكائنات العدوانية، والحيوانات المفترسة، والشخصيات التي تمثل القوة والسلطة يمكن تفسيرها تفسيراً رمزياً، وتحليلياً.

كما يجب أن نتذكر أن التفسير لا يتم على أساس استجابة واحدة، وإنما على أساس علاقة الاستجابة بالتنظيم الكلي للاستجابات.

ولكن يمكن أن تفيد كل استجابة في اقتراح فروض أو تفسيرات محتملة عن شخصية الفرد. وهنا ينبغي على الفاحص التحقق من هذه الفروض ومراجعتها في ضوء التفسيرات التي تعتمد على الاستجابات الأخرى، والتفسيرات التي تعتمد على النمط العام للاستجابات، وكذلك التفسيرات التي تعتمد على سلوك المفحوص أثناء استجابته لاختبار رورشاخ. وبالنسبة لهذه النقطة الاخيرة، فإن الفاحص يقوم بتسجيل ملاحظاته عن سلوك المفحوص أثناء استجابته، ويستخدم هذه الملاحظات كمصدر إضافي للبيانات عند تفسيره للاختبار. فالمفحوص الذي يطلب دائماً العون والإرشاد من الفاحص، ربما يفسر سلوكه هذا بأنه سلوك شخص غير مستقل.

والمفحوص الذي يبدو متوتراً، أو يسأل اسئلة مراوغة، ثم ينظر إلى الوجه الآخر للبطاقة، ربما يفسر هذا السلوك بأنه سلوك شخص متشكك، أو ربما يكون مصاباً بالبارانويا.

ويعتمد تفسير أي استجابة على مجموعة من خصائص هذه الاستجابة وهي: المحددات المستخدمة، وخصائص المدرك الحسي، والمحتوى، وما يقترن بالسلوك أثناء تقرير المدرك. كما يعتمد التفسير أو التقرير بوجه عام على تحليل كمي، وكيفي للاستجابات.

ثبات وصدق اختبار رورشاخ:

إن تنوع وثراء استجابات اختبار رورشاخ التي تستخدم في تقدير معظم أو جميع أبعاد شخصية الفرد يؤدي إلى مشكلات تتعلق بتقدير ثبات وصدق الاختبار.

فمعظم طرق تقدير ثبات الاختبارات السيكومترية، مثل : اختبارات ومقاييس الجوانب المعرفية، واستبيانات الشخصية تكون مناسبة لأن هذه الاختبارات والمقاييس تؤدي إلى نتائج كمية.

أما في حالة الاختبارات الإسقاطية مثل اختبار رورشاخ وغيره، فإننا لا نستطيع الرجوع إلى جداول معايير Norms Tables لمعرفة الدلالة المباشرة لأقسام الدرجات كما هو الحال في الاختبارات المقننة. ولذلك فإن إحدى طرق تقدير ثبات اختبار رورشاخ هو تحديد مدى اتفاق مجموعة من المحكمين في تقدير درجات الاختبار. ولكن هذه الطريقة غير شائعة الاستخدام بسبب اختلاف نظم تقدير الدرجات، كما أنها لا تقدم دليلاً على ثبات تفسيرات الدرجات. ويؤدي اختلاف طبيعة نظام تقدير درجات الاختبار إلى عدة مشكلات في تقدير ثبات الدرجات بطريقة اعادة الاختبار، أو بطريقة التجزئة النصفية. ومن بين هذه المشكلات صعوبة تحديد ما إذا كان عدم استقرار الدرجات يرجع إلى عدم استقرار طريقة القياس، أو إلى التغير المؤقت الذي يحدث في التنظيم الانفعالي للمفحوص. وتختلف بطاقات اختبار رورشاخ عن مفردات الاختبار السيكومتري، مما يجعل من الصعب تجديد تكافؤها. كما أن عدد استجابات الأفراد لبقع الحبر يختلف من بطاقة إلى أخرى. كل هذا يؤدي إلى صعوبات في تقسيم البطاقات إلى نصفين متكافئين.

والخلاصة أن عدم وجود أدلة كافية تؤكد ثبات اختبار رورشاخ يجعل من الصعب معرفة

ما إذا كانت الاستجابات غير مستقرة حقيقة – أي غير ثابتة – أم أنه يصعب القيام بدراسات مناسبة لتقدير ثبات تفسير هذا الاختبار.

أما التحقق من صدق اختبار رورشاخ، فإنه يعد أمراً بالغ التعقيد، إذ تزخر الأدبيات السيكلوجية بدراسات متناقضة في هذا الشأن، كما لا يوجد اتفاق بينها حول أفضل طريقة لتقدير صدق هذا الاختبار.

وإحدى طرق تقدير صدق الاختبار هو التحقق من مدى مضاهاة التفسيرات الناتجة عن الاختبار بالبيانات الناتجة من مصادر أخرى. فقد بينت بعض الدراسات وجود اتفاق بين تفسيرات اختبار رورشاخ، والتقرير أو الوصف الكلينيكي، بينما وجد آخرون اتفاقاً ضئيلاً بين تفسيرات الاختبار والتفسيرات التي تعتمد على الاختبارات الإسقاطية الأخرى (Lindzey, 1959) .

ويرى بعض الباحثين أن الطرق المعروفة لتقدير صدق الاختبارات لا يمكن تطبيقها على اختبار رورشاخ. إذ يصعب الاتفاق على المحك الذي يمكن استخدامه في التحقق من تفسيرات هذا الاختبار للمستويات العميقة أي اللاشعورية للسلوك.

(2) اختبار تفهم الموضوع The Thematic Apperception Test (TAT):

تختلف الاختبارات الإسقاطية في مدى تحديد وانتظام بنية كل منها. فالمثيرات التي تُكون اختبار الرورشاخ تعد اكثرها غموضاً وعدم انتظاماً، وتسمح بتنوع وثراء الاستجابات.

أما المثيرات التي يشتمل عليها اختبار تفهم الموضوع (TAT) فهي أكثر انتظاماً، والاستجابات تكون أكثر تحديداً من اختبار رورشاخ. ويتكون هذا الاختبار الذي أعده مورجان Morgan ، وموراي Murray عام 1935 من (30) صورة كل منها مرسومة على بطاقة منفصلة، وبطاقة بيضاء. وتختلف كل صورة من حيث محتواها ودرجة غموضها. وتتباين هذه الصور في درجة انتظامها وبنيتها، إذ تتراوح بين الانتظام التام أو البنية المحددة بدرجة كبيرة إلى البنية غير المحددة على الإطلاق كما في البطاقة البيضاء.

ويوضح الشكل (8 - 4) التالي إحدى هذه الصور:



شكل :(8 - 4) يوضح بطاقة صورة مماثلة للصور في اختبار تفهم الموضوع

ويمكن تطبيق الاختبار على الراشدين والمراهقين، وتُعرض البطاقات على الفرد واحدة تلو الاخرى مصحوبة بتعليمات بسيطة. ويُطبق الاختبار عادة على مرتين أو أكثر، وتُستخدم فقط (20) بطاقة مع أي فرد. ويُطلب من الفرد الذي يختبر أن يكون قصة عما حدث في الصورة، وعما يحدث فيها، وما سوف يمكن أن يحدث. ويجب أيضاً أن تتضمن القصة تفاصيلاً عن الشخصيات التي في الصورة، وأنشطتهم، وتفكيرهم، ومشاعرهم.

ويعتمد تقدير الدرجات على النظام المستخدم، إلا أن معظم أنظمة التقدير تتضمن التعرف على بطل أو بطلة القصة (أي البطل الذي يتطابق معه الفرد المختبر، أو يُسقط عليه شخصيته)، وتحديد حاجاته أو حاجاتها، وتحليل الضغوط أو القوى الاجتماعية التي تؤثر عليه أو عليها. وهذا النظام التقديري مستمد مباشرة من نظرية موراي في الشخصية.

ومما هو جدير بالذكر أن هناك بطاقات خاصة يمكن استخدامها مع الأطفال، ويسمى "اختبار تفهم الموضوع للأطفال Children Apperception Test (CAT)".

وتوجد مجموعات أخرى من البطاقات لقياس بعض متغيرات الشخصية، مثل البطاقات المعدلة لاختبار تفهم الموضوع، والتي استخدمها ماكليلاند (McClelland, 1953) لدراسة دافعية الإنجاز Achievement Motivation .

ويختلف اختبار تفهم الموضوع عن اختبار رورشاخ في أنه يُستخدم في الكشف عن جوانب الشخصية، مثل: الدوافع، والحاجات، والعواطف، والصراعات، والخيالات. بينما يستخدم اختبار رورشاخ في الكشف عن بناء وتنظيم الشخصية، ولا يهتم على وجه الخصوص بالمحتوى. فليس من المهم كثيراً في هذا الاختبار أن يشير المفحوص إلى جزء معين من بقعة الحبر على أنه ثعبان، أو حبل، ولكن المهم هو مدى معقولية المدرك، وهل تحدد عن طريق شكل أو لون هذه المنطقة من البقعة.

ويعد تقدير صدق اختبار تفهم الموضوع من الأمور المعقدة ايضاً كما هو الحال في اختبار رورشاخ. فهل القصص تعكس حقيقة شخصية الفرد المختبر، أم أنها مجرد ردود فعل نمطية للموقف المصور؟ وحتى إذا افترضنا أن القصص تعكس حقيقة شخصية الفرد، فهل افتراض أن الفرد يتطابق مع البطل أو البطلة، وأن مشكلات البطل أو البطلة، والضغوط المؤثرة على اي منهما تعكس بالفعل جوانب حياة الفرد المختبر، يعد افتراضاً صحيحاً؟

ومما يزيد من تعقد المشكلة عدم الاتفاق على طريقة واحدة لتقدير الدرجات، ونظام تفسيرها. كما أن تقدير الصدق الخارجي للاختبار يعد من المشكلات التي لم تحسم بعد.

وعلى أية حال، فإن اختبار رورشاخ، واختبار تفهم الموضوع يعد كلاً منهما أداة كلينيكية تتطلب أخصائياً متمرساً لاستخدامها، وتفسير استجابات المختبرين. ويفضل كثير من علماء النفس الكلينيكي استخدام الاختبارين معاً للتوصل إلى فهم متكامل للشخصية.

وينبغي الاشارة إلى أن هناك اختبارات إسقاطية كثيرة أخرى تعتمد على طريقة التداعي، مثل: اختبار تداعي الكلمات Word Association Test الذي أعده كنت، ورسانوف Kent & Rosanoff، وطريقة الترتيب Ordering مثل: اختبار زوندي Szondi، وطريقة التعبير مثل: تمثيل الأدوار، والسيكودراما، واللعب بالدمى، وطريقة الرسم، مثل اختبار رسم الرجل Draw A - Person Test، واختبار تذكر وإنتاج تصميمات هندسية معينة من الذاكرة بعد عرض تسعة أشكال هندسية واحداً تلو الآخر كل على بطاقة منفصلة، ويسمى اختبار بندر، وجستلت Bender & Gestalt Test، وغيرها.

تعقيب عام على المقاييس الإسقاطية للشخصية:

يتضح مما سبق أن من أهم مميزات المقاييس الإسقاطية للشخصية حرية الاستجابة المسموح بها للفرد المختبر، وحرية الأخصائي الكلينيكي في سبْر وتوضيح ما يقصده الفرد. ومعظم هذه المقاييس تكون غامضة الهدف، وبذلك يصعب على الفرد المختبر تزوير استجاباته، لأن طرق تقدير الدرجات، وتفسير الاستجابات، تكون غير مألوفة لديه. كما أن معظمها لا يحتاج إلى مهارة في القراءة ويمكن أن تطبق على الأميين والأطفال الصغار. إلا أن هذه المقاييس يشوبها كثيراً من العيوب، فهي تطبق بطريقة فردية، ولذلك فهي مكلفة في التطبيق، وفي تقدير الدرجات والتفسير. ومما يزيد الأمر تعقيداً عدم وجود أدلة كافية على جدوى استخدامها، وثباتها، وصدقها.

ويجب أن يقوم بتطبيق هذه المقاييس، وبخاصة تلك التي تكون بنيتها غير محددة، مثل: اختبار رورشاخ، واختبار تفهم الموضوع، والسيكودراما، واللعب، واختبارات الرسم، أخصائيون متمرسون على كيفية استخدام هذه الاختبارات، وتفسير نتائجها.

الفصل التاسع

اختيار وتقييم الاختبارات والمقاييس

- مقدمة
- اعتبارات عملية عند اختيار الاختبارات والمقاييس
- اعتبارات فنية عند اختيار الاختبارات والمقاييس
- دليل الاختبار
- مصادر الحصول على معلومات عن الاختبارات
- المقاييس المنشورة

مقدمة:

تناولنا في الفصول السابقة كيفية قياس السمات أو الخصائص الإنسانية المتعددة المتمثلة في التحصيل، والذكاء العام، والاستعدادات، والشخصية، واوضحنا طبيعة الاختبارات والمقاييس المناسبة لها، وأنواعها المختلفة، وخصائصها السيكومترية، وكيفية تفسير درجاتها. ولعل التساؤل المهم الذي يبرز الآن يتعلق بكيفية اختيار الاختبار أو المقياس المناسب لغرض أو اغراض معينة من بين العديد من الاختبارات المنشورة والمتوافرة، بحيث يمكن الاستناد إلى نتائجها في اتخاذ القرارات المتعلقة بالأفراد، أو الجماعات، أو البرامج بدرجة عالية من الثقة. وكذلك الأسس التي يمكن الإستناد إليها في تقييم هذه الاختبارات والمقاييس التي يتم اختيارها للغرض أو الاغراض المعينة.

لذلك سوف يتناول هذا الفصل بعض العوامل أو الاعتبارات العملية والفنية التي ينبغي مراعاتها عند اختيار الاختبارات والمقاييس لاغراض معينة، والمعايير اللازم توافرها في أدلة الاختبارات، ومصادر الحصول على معلومات عن الاختبارات والمقاييس المنشورة.

اعتبارات عملية عند اختيار الاختبارات والمقاييس:

عند اختيار اختبار أو مقياس لاستخدامه في غرض معين، ينبغي مراعاة بعض الاعتبارات العملية التي تؤثر في هذا الاختيار، لعل أهمها ما يلي:

(1) **السمة المراد قياسها، والغرض من القياس:** فقد سبق ان اوضحنا في كثير من المواضع أن تحديد السمة المراد قياسها يعد خطوة اساسية توجه عملية اختيار أداة القياس المناسبة لهذه السمة. فإذا كنا نود قياس الذكاء العام، فإننا سوف نقتصر على الاختبارات والمقاييس العقلية الفردية أو الجماعية التي تقيس العامل العام (G)، واذا كنا نود قياس الشخصية، فإننا ينبغي تحديد السمات التي نهتم بها أكثر من غيرها، لأن ذلك يجعل اختيار استبيان الشخصية المناسب أكثر تحديداً، وهكذا. كما ينبغي مراعاة الغرض من القياس سواء كان من أجل الانتقاء أو التسكين، أو التصنيف، أو

التشخيص مثلاً. وهذا التحديد يعد أمراً ضرورياً لأنه يوجه عملية الاختيار الوجهة الصحيحة، فالاختبار الذي يهدف للانتقاء يختلف في محتواه وتفسير درجاته عن محتوى الاختبار الذي يُصمم من اجل التشخيص، حيث يشتمل في هذه الحالة على مفردات تقيس مختلف عناصر المحتوى أو مجالاته الفرعية، ويمدنا بدرجات لكل من هذه المجالات.

(2) طبيعة المجموعة المستهدفة من الأفراد: فقد لاحظنا في الفصول الأربعة السابقة أن اختبارات الراشدين تختلف عن اختبارات الأطفال، والأميين مثلاً ؛ فالاختبارات تُصمم عادة بحيث تناسب فئة أو فئات معينة، وأعماراً مختلفة. لذلك فإن تحديد طبيعة المجموعة المراد اختبارها يُيسر عملية اختيار الاختبار أو الاختبارات المناسبة لها. وعندما يتباين أداء الأفراد بتباين خصائص معينة، مثل: العمر، والتعليم، والمستوى الاجتماعي الاقتصادي، وغير ذلك، فانه ينبغي مراعاة هذه الخصائص عند اختيار الاختبار أو المقياس.

(3) كلفة العملية الاختبارية: فتطبيق الاختبارات التحصيلية المقننة على نطاق واسع في المدارس – كما هو الحال في كثير من الدول – يعد عملية مكلفة. ولا تقتصر الكلفة على شراء كراسة الاختبار، واوراق الاجابة، وإنما أيضاً تتضمن تصحيح الاختبار، وتقرير نتائجه. وتزداد الكلفة عندما تُستخدم كراسة الاختبار مرة واحدة، حيث تتطلب الإجابة وضع علامات على كراسة الاختبار مباشرة. كما ان اعتماد الاختبار على أجهزة معينة أو قاعات خاصة ربما يضيف مزيداً من كلفة تطبيق الاختبار. غير انه لا ينبغي التغاضي عن استخدام اختبار جيد بسبب كلفته العالية.

(4) زمن تطبيق الاختبار: فقد سبق أن اوضحنا أن بعض الاختبارات، وبخاصة اختبارات الذكاء الفردية، وبعض استبيانات ومقاييس الشخصية يستغرق تطبيقها كثيراً من الوقت. وهناك مواقف عملية تتطلب تطبيق اختبار معين في زمن لا يزيد عن ساعة واحدة مثلاً، فعندئذ ينبغي مراعاة ذلك عند اختيار الاختبار أو المقياس. وعلى الرغم من أن الاختبار القصير يكون أقل كلفة، ويستغرق تطبيقه وتصحيحه وقتاً محدوداً، إلا أن ثبات درجاته يكون منخفضاً لقلة عدد مفرداته. لذلك ينبغي مراعاة التوازن بين زمن تطبيق الاختبار، وثبات درجاته، عند اختيار الاختبار المناسب.

(5) سهولة تطبيق الاختبار: عند اختيار اختبار أو مقياس معين ينبغي مراعاة مدى سهولة التطبيق، وكفاية التعليمات الموضحة بدليل الاختبار. فكلما كانت تعليمات التطبيق *كاملة، أمكن جعل عملية التطبيق منتظمة. فالتعليمات التفصيلية لكل من الفرد المختبر،* والقائم بتطبيق الاختبار، تُيسر العملية الاختبارية. وينبغي أن تشتمل هذه التعليمات على معلومات تتعلق بزمن الاختبار إذا كان من النوع الذي يتطلب السرعة في الإجابة، وذلك لكل قسم من أقسامه، وتنظيم المقاعد، وكيفية المراقبة أثناء أداء الاختبار، ونوع القلم الذي يُستخدم في وضع العلامات في ورقة الاجابة، وطرق تصحيح اخطاء الإجابة سواء بالشطب أو بوضع علامة (x) على الإجابة المراد تصحيحها، ومستوى الإضاءة، ودرجة حرارة الغرفة ومتطلبات تهويتها، والإجراءات التي تتبع إذا خالف فرد شروط العملية الاختبارية، وكذلك الاجراءات التي تُتبع في حالات الطوارىء، وطريقة توزيع اوراق أو كراسات اجابة الاختبار وجمعها، وطرق تصحيح وتقدير درجات هذه الاوراق أو الكراسات. ويجب على مستخدم الاختبار قراءة دليل الاختبار بعناية ويتخذ قراراً فيما يتعلق بكفاية تعليمات تطبيقه.

(6) ظروف تطبيق الاختبار: فالظروف الفيزيائية للعملية الاختبارية تعد من الأمور التي ينبغي مراعاتها، مثل: تنظيم مقاعد غرفة الاختبار، وإضاءة الغرفة، ونظام التهوية، والحرارة، وغير ذلك. فهذه الظروف تؤثر بدرجة ما في درجة الفرد في الاختبار. وينبغي أن يُختبر جميع الأفراد في ظروف تماثل بدرجة كبيرة الظروف التي كانت متواجدة وقت تقنين الاختبار.

(7) الجوانب النفسية للمختبرين: فالاتجاه الإيجابي للأفراد نحو الاختبار والعملية الاختبارية، يعد من الاعتبارات المهمة. فالاختبار الذي لا يراعى عند تطبيقه الجوانب النفسية للمختبرين يؤدي إلى نتائج غير موثوق بها. فقلق الاختبار مثلاً يؤثر في الأداء الاقصى للمختبرين.وهذا يلقى دوراً على القائم بتطبيق الاختبار لتهيئة الأفراد للاختبار من الناحية الانفعالية، وزيادة دافعيتهم لبذل أقصى جهدهم دون قلق. فإذا أدرك المختبرون الفائدة التي تعود عليهم من النتائج الدقيقة للاختبار، فإن ذلك سوف ينعكس إيجابياً على تهيئتهم التهيئة الانفعالية المناسبة.

(8) صيغة الاختبار: فالاختبار الذي تكون صيغته، وطريقة تنظيمه متناسبة مع طبيعة المجموعة المستهدفة، بحيث يراعي مستوى نضج أفراد المجموعة، وعدم ازدحام الاختبار بالمفردات، ووضوح الصور والأشكال، وضم جميع التعليمات في مكان واحد، كل ذلك يزيد من دافعية المختبرين لبذل أقصى جهدهم في الاختبارات المقننة. فمثلاً دليل اختبار الاستعدادات الفارقة (DAT) الذي سبق أن اوضحناه في الفصل الخامس يفصل بوضوح بين تعليمات تطبيق الاختبار، وطريقة تصحيحه، ومعاييره، ويضع عناوين جانبية واضحة لكل من هذه العناصر.

(9) طريقة التصحيح: توجد انواع مختلفة من طرق تصحيح الاختبارات والمقاييس؛ فالاختبارات التي تتطلب ورقة إجابة منفصلة يمكن تصحيحها يدوياً أو آلياً باستخدام الماسح الضوئي Optical Scanner . وقد أصبحت هذه الماسحات شائعة الاستخدام في الآونة الأخيرة لتصحيح آلاف من اوراق إجابة الاختبارات التي تشتمل على مفردات الاختيار من متعدد، أو الصواب أو الخطأ، وذلك في وقت قصير. كما أن بعض الاختبارات يتم تصحيحها، وتقرير نتائجها آلياً باستخدام الحاسوب.

(10) توافر صيغ متكافئة من الاختبار : فهذه الصيغ المتكافئة Alternate Forms تساعد في إعادة اختبار الأفراد في اوقات مختلفة دون الحاجة إلى تطبيق الاختبار نفسه، مما يحد من مشكلة سرية الاختبارات. فاستخدام صيغة مكافئة يجعل من الصعب تذكر مفردات الصيغة الاولى التي سبق تطبيقها على الفرد نفسه. لذلك ينبغي مراعاة الهدف من العملية الاختبارية، واتخاذ قرار عما إذا كانت تستلزم توافر صيغتين متكافئتين من الاختبارنفسه، وذلك يوجه عملية اختيار الاختبار المناسب.

(11) سهولة تفسير درجات الاختبار : ينبغي أن يتضمن دليل الاختبار معلومات تُيسِّر تفسير درجات الاختبار. وهذا يتطلب تقديم معلومات تتعلق بمعايير الاختبار Test Norms، ومزاياها، وعيوبها، وكذلك أمثلة لتفسيرات غير صحيحة من أجل العمل علي تجنبها، وأمثلة أخرى لتفسيرات مناسبة. وكثير من التفسيرات غير الصحيحة ربما ترجع إلى أمور فنية تتعلق بثبات الدرجات، وصدق الاختبار، كما سيتضح بعد قليل.

اعتبارات فنية عند اختيار الاختبارات والمقاييس :

إن مراعاة الاعتبارات العملية السابقة عند اختيار الاختبارات والمقاييس ليست كافية لاتخاذ قرار في هذا الشأن. فإذا كان هناك اختباران أو أكثر يختلفان اختلافاً كبيراً في هذه الاعتبارات العملية ينبغي التفكير جيداً في المزايا النسبية لكل منها إذا استُخدم في الموقف الاختبارى المعين. إذ ليس من الحكمة أحياناً اختيار اختبار أو مقياس على أساس سهولة تطبيقه، وتصحيحه مثلاً، في حين أنه يفتقر إلى الصدق والثبات، مما يجعل نتائج الاختبار غير موثوق بها، وبذلك يكون مضيعة للوقت والجهد.

لذلك ينبغي عند اختيار الاختبارات والمقاييس مراعاة اعتبارات فنية أيضاً فيما يتعلق بالأدلة المقارنة لكل من الصدق والثبات، وكفاية المعايير. وفيما يلي توضيحاً لكل من هذه الخصائص:

(1) صدق الاختبار : يُعد الصدق Validity وكيفية تقديره من الاعتبارات الفنية المهمة التي ينبغي التحقق منها عند اختيار اختبار لغرض معين. وقد سبق أن اوضحت مفهوم الصدق وطرق تقديره في الفصل الرابع. فالاختبارات المقننة والاستبيانات ينبغي أن تكون صادقة فيما تقيسه، بما يؤدي إلى الثقة في استخداماتها لأغراض معينة.

فالاختبارات التحصيلية مثلاً ينبغي أن تتميز بصدق محتواها، بحيث تقيس التحصيل في مجال دراسى معين، وتشتمل على عينة ممثلة لنطاق المهارات التي يتضمنها هذا المجال. وهذا يعنى وجود علاقة بين التعلم المراد تقييمه، والمواد الاختبارية التي يُفترض أنها تقيس نواتج هذا التعلم. لذلك فإن توافر بيانات تتعلق بصدق الاختبار يعد أمراً ضرورياً لتقديم أدلة عن أنواع الصدق المناسبة لاستخدامات الاختبار. وهذه البيانات تكون مذكورة عادة في دليل فنى للاختبار Technical Manual. لذلك ينبغي تقييم صدق الاختبار في علاقته بالقرارات أو الأغراض التي سوف يُستخدم فيها الاختبار.

(2) ثبات درجات الاختبار : يعد الثبات Reliability خاصية تلى الصدق في أهميتها، وينبغي أخذها بعين الاعتبار عند اختيار اختبار أو مقياس معين. فالغرض من استخدام نتائج الاختبار سوف تحدد مستوى الثبات المقبول. فإذا كان الغرض من هذا الاستخدام تقييم الفرد، ينبغي أن يكون معامل الثبات أعلى بكثير مما لو كان الغرض

هو التقييم الجماعي. ففي الحالة الاولى ربما يُفضل أن تقترب قيمة معامل الثبات من (0.90)، بينما يكفى في الحالة الثانية أن تكون هذه القيمة (0.50) أو أكبر قليلاً (Cronbach, 170; Seibel, 1968) . وبالطبع ينبغي أن يبين دليل الاختبار كيفية تقدير ثبات درجات الاختبار. فمعامل الثبات يعد مهماً في المقارنة بين الاختبارات لأنه يكون مؤشراً لاتساق ما يقيسه الاختبار. وكذلك يرتبط الخطأ المعياري للقياس Standard Error of Measurement بمفهوم الثبات، وذلك لأنه يعد مقياساً! إحصائياً مهماً إذا أردنا أن تُفسر درجات الاختبار بطريقة واحدة.

وكلما كانت قيمة معامل الثبات مرتفعة، والخطأ المعيارى للقياس منخفضاً، فإن الاختبارات يكون لها ثبات أكبر .

لذلك ينبغي معرفة إجراءات تقدر ثبات درجات الاختبار، وحجم العينة، وقيمة معامل الثبات الذي يتحدد في دليل الاختبار، وكذلك تقدير الخطأ المعياري للقياس، وتحديد مصادر أخطاء القياس .

(3) معايير الاختبار: وتتعلق بتفسير درجات الاختبار، وتعتمد هذه المعايير Norms على حجم عينة التقنين الذي ينبغي أن يكون كبيراً نسبياً، لكي يزداد احتمال تمثيلها للمجتمع المستهدف. وينبغي أن تراعي هذه العينة المناطق الجغرافية المختلفة، والمستويات الاقتصادية والاجتماعية، والنوع، والعمر، والفرق الدراسية، والخلفية الثقافية، وغير ذلك. وهذا يساعد في التوصل إلى استنتاجات صادقة من جداول المعايير عندما تُطبق على مجموعة أخرى مماثلة لعينة التقنين. وعادة يُوضح الدليل الفني للاختبار خصائص عينة التقنين، وتاريخ جمع البيانات المتعلقة بالمعايير، ونوع الدرجات المحوَّلة المستخدمة، أي: الدرجات التائية، أو الرتب المئينية، أو التساعيات المعيارية، أو نسب الذكاء الانحرافية، أو غير ذلك .

وينبغي عند اختياراختبار أو مقياس معين مراعاة أن تكون المعايير حديثة بحيث تواكب التغيرات المتسارعة في مجالات التعليم والمهن المختلفة، وأن تكون قابلة للمقارنة بين الاختبارات المختلفة، وبخاصة عندما تُقارن الصفحات النفسية (البروفيلات) لأغراض التشخيص. كما يجب وصف هذه المعايير وصفاً تفصيلياً يوضح طبيعة عينة التقنين، وإجراءات الحصول على جداول الدرجات المحوَّلة. وينبغي أن يشتمل ذلك على أسلوب المعاينة،

وعدد أفراد العينة المستخدمة، وتوزيعها، وخصائصها الديموغرافية، وتاريخ الحصول على هذه المعايير. فمعرفة جميع هذه المعلومات يساعد في الحكم على مدى ملاءمة معايير الاختبار للغرض أو الأغراض المعينة .

دليل الاختبار The Test Manual :

يُعد دليل الاختبار هو المصدر الرئيس للمعلومات المتعلقة بجودة ونوعية الاختبار أو المقياس المنشور، وعادة يكون ملحقاً به كتيباً يتضمن البيانات والخصائص السيكومترية أو الفنية للاختبار.

وينبغي أن يكون دليل الاختبار واضحاً بدرجة كافية بحيث يمكن لمستخدمي الاختبار مهما اختلفت مؤهلاتهم فهم محتوياته. كما ينبغي أن تكون المعلومات التي يشتمل عليها دقيقة بحيث تفي باحتياجات المتخصصين في بحوث الاختبارات والمقاييس.

ونظراً لأهمية أدلة الاختبارات أعدت لجنة مشتركة من أعضاء الجمعية الأمريكية لعلم النفس (APA)، والجمعية الأمريكية للبحث التربوى (AERA)، والمجلس القومى الأمريكى للقياس التربوي (NCME) مجموعة من المعايير Standards يمكن الاسترشاد بها عند إعداد الاختبارات والمقاييس التربوية والنفسية ، وأدلتها. وهذه المعايير تتجدد باستمرار نظراً للتطورات المتسارعة في علم القياس النفسي والتربوي. ولكن ليس من الضرورى من الوجهة القانونية أن يلتزم ناشرو الاختبارات بهذه المعايير، لذلك ربما نجد اختباراً أو مقياساً منشوراً يصحبه دليل غير جيد.

غير إنه إذا استوفى دليل الاختبار هذه المعايير، فإنه يكون مشتملاً على المعلومات الضرورية التي تساعد في اختيار الاختبار المناسب لاستخدام معين. وقد صُنِّفت هذه المعايير في العناصر التالية (APA, 1999) :

(أ) نشر المعلومات :

(1) عند نشر اختبار للاستخدام العملى، ينبغي أن يصاحبه دليل أو مجموعة من الأدلة التي توفر لمستخدمي المعلومات التي يحتاجونها لكي يتمكنوا من تفسير درجات الاختبار .

(2) ينبغي تحديث الاختبار ودليله أو أدلته بصورة منتظمة، نظراً لأن المعلومات المتعلقة بالاختبار تتقادم سريعاً، وبخاصة مع التغير المستمر في الظروف الاجتماعية، ومواصفات الأعمال، والضغوط التربوية. وهذا يؤدي إلي تغير تفسير درجات بعض الاختبارات. وربما ينبغي إيقاف استخدام الاختبار إذا لم تُجر مراجعة على دليل الاختبار في غضون (15) عاماً منذ نشره .

(ب) تفسير درجات اختبار :

(1) ينبغي أن يساعد الاختبار، ودليله، ومواده المصاحبة مستخدمي الاختبار في تفسير درجاته تفسيراً صحيحاً له معنى .

(2) ينبغي أن يحدد دليل الاختبار بوضوح أغراض استخدام الاختبار وتطبيقاته التي يُوصى بهاً .

(3) ينبغي أن يحدد دليل الاختبار المؤهلات المطلوبة لتطبيق الاختبار، وتفسير درجاته تفسيراً مناسباً .

(جـ) صدق الاختبار :

(1) ينبغي أن يوضح دليل الاختبار صدق الاختبار لكل نوع من الاستدلالات التي يُوصي بها.

(2) إذا صُمِّم الاختبار بدرجة أساسية لمعاونة مستخدمى الاختبار في إصدار أحكام ذات طبيعة تنبؤية، فإنه ينبغي تقدير الصدق التنبؤي للاختبار بعناية .

(3) إذا صُمِّم الاختبار بدرجة أساسية لتقديم معلومات تتعلق بالتحصيل السابق للفرد، فإنه يجب تقدير صدق محتوى الاختبار، لأن هذا النوع من الصدق يكون أكثر ملاءمة في هذه الحالة.

(4) إذا تم تقدير الصدق التنبؤي أو التلازمي للاختبار، فإنه ينبغي وصف المقاييس المحكة Measures Criterion التي استُخدمت في هذا الشأن بعناية. وكذلك تقرير معلومات تصف العينة المستخدمة في دراسة الصدق، وظروف إجراء الاختبار. وهذا يساعد مستخدمي الاختبار في الحكم على مدى ملاءمة البيانات المتعلقة بصدقه للأغراض المناسبة لهم .

(5) ينبغي عرض نتائج التحليل الإحصائى الذى استُخدم في تقدير الصدق المرتبط بمحك بطريقة واضحة ومفهومة .

(6) إذا اقترح مؤلف الاختبار تفسير درجاته كمقياس لمتغير نظري (قدرة، أو سمة، أو اتجاه معين)، ينبغي تحديد التفسير المقترح تحديداً كاملاً. فمثلاً إذا افتُرض أن الاختبار يقيس تكويناً فرضياً Construct، مثل القلق، فإنه ينبغي أن يُميز دليل الاختبار بين تعريف القلق المُستخدم، والتعريفات المحتملة الأخرى .

(د) ثبات درجات الاختبار:

(1) ينبغي أن يُقدم دليل الاختبار معلومات عن ثبات درجاته، بحيث تسمح لمستخدمي الاختبار الحكم على ما إذا كانت الدرجات يمكن الاعتماد عليها بدرجة كافية للغرض الذي سوف يُستخدم الاختبار من أجله. وإذا لم يتم جمع أي معلومات أو أدلة ضرورية، فإنه ينبغي ذكر المعلومات الناقصة .

(2) إذا قدَّم دليل الاختبار تقديرات الثبات، أو تقديرت أخطاء القياس، فإنه ينبغي تحديد الإجراءات التي استُخدمت في الحصول على هذه التقديرات. وكذك وصف عينة الأفراد الذين طُبِّق عليهم الاختبار، وخصائص العينة، مثل: مستوى نُضج الأفراد، وتباين قدرتهم، واتجاههم نحو الاختبار، وغير ذلك، حيث إن هذه المتغيرات ربما تؤثر في هذه التقديرات .

(3) إذا تم الحصول على معاملات الثبات من مجموعة غير متجانسة من الأفراد، وأراد مستخدم الاختبار تطبيقه على مجموعة متجانسة، فإنه يجب ألا يتوقع الحصول على نفس قيمة معامل الثبات الذى حدَّده دليل الاختبار .

(4) إذا كان للاختبار أكثر من صيغة واحدة، فإنه ينبغي أن يُقدم دليل الاختبار بيانات فيما يتعلق بتكافؤ هذه الصيغ الأخرى .

(5) إذا صُمِّم الاختبار لقياس سمة عامة متجانسة، فإنه ينبغي أن يُقدِّم دليل الاختبار ما يُوضح الاتساق الداخلى لدرجاته .

(6) إذا استُخدم الاختبار في أغراض التنبؤ، أو مقارنة أداء الفرد عبر الزمن، فإن معامل

الاستقرار بطريقة إعادة الاختبار يكون ضرورياً، ويتطلب تقديراً لقيمته، وتقرير ذلك في دليل الاختبار .

(هـ) تطبيق الاختبار، وتصحيحه :

(1) ينبغي أن تكون تعليمات الاختبار واضحة بدرجة كافية في دليل الاختبار بحيث يمكن اتباعها، وتكرارها كما في عينة التقنين.

(2) ينبغي أن تكون إجراءات تصحيح الاختبار محددة وتفصيلية وواضحة في دليل الاختبار للتقليل بقدر الإمكان من أخطاء التصحيح .

(و) الموازين، والمعايير:

(1) الموازين المستخدمة في تقرير نتائج الاختبار Scales ينبغي أن توصف بعناية في دليل الاختبار لزيادة احتمال التفسير الصحيح للدرجات.

(2) إذا أُجريت مراجعة لموازين الاختبار، أو أُضيفت موازين جديدة، أو أي تعديلات أخرى، فإن دليل الاختبار المنقَّح ينبغي أن يُقدم الجداول التي توضح التكافؤ بين الموازين الأصلية والموازين الجديدة .

(3) معايير الاختبار Test Norms ينبغي أن توصف بعناية، بحيث يمكن لمستخدمي الاختبار تحديد ما إذا كانت عينة التقنين مماثلة للمجموعة التي سوف يُستخدم معها الاختبار .

وهذا يتطلب أن يشتمل دليل الاختبار على وصف دقيق للمجتمع المستهدف الذى يُقارن به الأفراد المختبرون، ومتوسطات درجات الاختبار، وانحرافاتها المعيارية، وكذلك الخصائص الديموغرافية لهذا المجتمع .

(4) ينبغي أن يشتمل دليل الاختبار على جداول المعايير التي تحتوى على الدرجات الخام في الاختبار، والدرجات المحوَّلة المناظرة لها، مثل : الدرجات المعيارية أو التائية، أو التساعيات المعيارية، أو الرتب المئينية التي تعكس توزيع الدرجات في مجموعة أو مجموعات معيارية مناسبة من الأفراد.

(5) ينبغي أن يشتمل دليل الاختبار على وصف متكامل لأى صفحة نفسية (بروفيل) يمكن تكوينها من بطارية تشتمل على عدد من الاختبارات الفرعية .

(6) ينبغي نشر البيانات المتعلقة بمعايير الاختبار في نفس وقت نشر الاختبار .

مصادر الحصول على معلومات عن الاختبارات والمقاييس المنشورة :

يعتمد اختيار الاختبار أو المقياس المناسب لغرض معين على معرفة مستخدمي الاختبارات مصادر الحصول على معلومات عن مختلف أنواع الاختبارات والمقاييس الأجنبية والعربية المنشورة. فالفرد الذى يود استخدام اختبارات معينة يحتاج إلى معرفة الاختبارات المتوافرة التي تساعده في الحصول على المعلومات التي تهمه، وآراء الخبراء في الاختبارات التي سوف يختارها، والبحوث التي أجريت حولها، والبيانات السيكومترية المتوافرة المتعلقة بصدقها وثبات درجاتها، ومجموعات الأفراد التي يناسبها اختبار معين، وهذه المعلومات تتطلب الرجوع إلى مصادر مختلفة متخصصة، سوف اوضح فيما يلى بعضاً منها :

(1) الكتاب السنوي للقياس العقلي : Buros' Mental Measurement Year book:

يُعد هذا الكتاب الدوري من أهم مصادر المعلومات فيما يتعلق بالاختبارات والمقاييس، وهو يصدر كل عام منذ عام 1938، وكان يُحرره اوسكار بيروس Oscar Buros حتى وفاته منذ عدة أعوام، ولا يزال يصدر سنوياً عن معهد بيروس للقياس العقلي بأمريكا، ويشتمل على مجموعة متجددة وشاملة من المصنفات، ومراجعات الاختبارات، والمراجع المتعلقة بالقياس والتقويم النفسي والتربوي. وقد صدر عن هذه المؤسسة حوالي تسع كتب سنوية حتى نهاية القرن الماضي. وتهدف هذه الكتب إلى تنمية الاتجاه الناقد نحو الاختبارات والعمليات الاختبارية من جانب ناشرى الاختبارات ومستخدميها، وتفعيل التواصل، وتحسين جودة ونوعية الاختبارات المنشورة. وكذلك تقديم معلومات تتعلق بالاختبارات المنشورة التي تتناول بناء الاختبارات المقاييس واستخداماتها وصدقها، والمراجع الجديدة المتعلقة بالعمليات الاختبارية، ومقالات ناقدة ممثلة للمُراجعات التي تُنشر في الدوريات المتخصصة حول هذه المراجع .

(2) اختبارات في مرحلة الطباعة Tests in Print :

وهي عبارة عن قائمة شاملة للاختبارات المستخدمة في المجالات النفسية، والتربوية،

والصناعية، ويقوم بنشرها أيضاً معهد بيروس للقياس العقلي. ويشتمل كل اختبار مذكور في هذه القائمة على معلومات تتعلق بعنوان الاختبار، والفرقة الدراسية المناسبة، وتاريخ النشر، وتعليق موجز حول الاختبار، وعدد ونوع الدرجات الناتجة، والمؤلف أو المؤلفين، والناشر، وإشارة إلى مراجعة الاختبار في الكتاب السنوى للقياس العقلى .

(3) منشورات مركز دراسات التقويم Center for the Study of Evaluation (CSE) :

يُصدر هذا المركز التابع لجامعة كاليفورنيا بأمريكا مجلدات تشتمل على دراسات منشورة تتعلق بتقييم اختبارات مختلفة. وتستند محكات التقييم إلى الصدق، والملاءمة، والواقعية، وجودة المعايير، وتقييم عام للاختبار .

(4) نشرة دورية تتعلق بالاختبارات News on Tests :

تصدر هذه النشرة الدورية عن مركز الخدمات الاختبارية Educational Testing Service (ETS) بولاية نيوجيرسي الأمريكية، وكان يُطلق عليها فيما سبق " نشرة تجميع الاختبارات Test Collection Bulletin ". ويحتفظ هذا المركز بمجموعات كبيرة من الاختبارات المنشورة وغير المنشورة التي يمكن الاطلاع على مسمياتها على موقع الإنترنت، أو على الأفلام المصغرة Microfiche عبر النظام ERIC.

كما تشتمل هذه النشرة على ملخصات موجزة للإضافات المستحدثة لمجموعة الاختبارات، ولكن لا توجد مراجعة لهذه الاختبارات كما في الكتاب السنوي للقياس العقلي .

(5) قوائم إصدارات الاختبارات Bibliographies of Tests :

توجد العديد من قوائم إصدارات الاختبارات التي تصف الاختبارات المنشورة وغير المنشورة. وهى قوائم لمجالات معينة تخصصية مثل: قياسات تتعلق بنمو الطفل، وتقييم مهارات القراءة، وغير ذلك. وكثير من هذه الإصدارات مذكورة في مجلدي (Baches, 1977; Aaronson, 1974) .

(6) دوريات القياس والتقويم Professional Journals :

توجد دوريات متعددة متخصصة في القياس والتقويم النفسي والتربوي، وتُجري مراجعات



للاختبارات التي يستخدمها المربون، والمشتغلون بعلم النفس. ويمكن الرجوع إلى الكتاب السنوى للقياس العقلي المذكور سابقاً للحصول على معلومات تتعلق بهذه الدوريات .

مؤسسات نشر الاختبارات : يمكن الحصول على معلومات تتعلق بالاختبارات والمقاييس النفسية والتربوية من ناشري الاختبارات في الولايات المتحدة الأمريكية بخاصة، حيث يمكن من خلال مواقعهم على الإنترنت معرفة أنواع الاختبارات والمقاييس، ووصفها، وكذلك من فهارس إصدارات هذه المؤسسات. وتشتمل هذه الفهارس على الإصدارات الحديثة للاختبارات المختلفة، والخدمات المتعلقة بتصحيحها، وكلفتها، وعينات من مفرداتها، وأدلتها، وتقارير فنية متعلقة بها .

وكذلك يمكن معرفة أسماء وأنواع الاختبارات والمقاييس التي يتم تعريبها وتقنينها في الوطن العربي من دور النشر الكبرى، والمكتبات المتخصصة ، ومواقعها على الإنترنت.

مراجع الكتاب

Adkins, D. (1974). Test Consruction: Development and Interpretation of Achievement Tests (2 nd ed.). Columbus: Charles Merill Publishing Company.

Aiken, L. (1973). Readings in Psychological and Educational Testing. Boston: Allyn and Bacon, Inc.

Aiken, L. (1976).Psychological Testing and Assessment, (2nd ed.) Boston: allyn and Bacon, Inc.

Aiken, L. (1987). Assessment of Intellectual Functioning. Boston: Allyn and Bacon, Inc.

Airasian, P. (1999). Classroom Assessment. N. Y. :McGraw Hill.

Allam, S. (1981) Minimum Measurement Competency Test: New Jersey: Educational Testing Service.

Allam, S. (1990). Systems and Methods of Evaluation and Examination in the Arab States: Present Status and Future Trends. Jordan: UNESCO, IPDAS.

Allport, G. (1963). Patterns and Growth in Personality. N.J.: Holt Reinehart & Winston.

American Psychological Association (1999). Standards for Educational and Psychological Testing. Washington, D.C.: Author.

Anastasi, A. (1988). Psychological Testing (6th ed.). N.Y: Macmillan.

Baker, T. (1977). A Directory of Information of Tests. Princeton, N.J.: ERIC Clearinghouse on Tests.

Bejar, I. (1974). Achievement Testing: Recent Advances. Calif: SAGE publications Inc.

Belanoff, P. & Dckson, M. (1991). Portfolios: Process and Product. Portsmouth, N.H: Boynton Cook Publishers.

Berk, R. (1986). Performance Assessment Methods and Applications. Baltimore MD: John Hopkins University Press.

Bertrand, A. & Cebula, J. (1980). Tests, Measurement, and Evaluation: A Developmental Approach. Mass.: Addision Wesly Publishing Company

Birenbaum, M. & Dochy, F (1996). Alternative Assessment of Achievement, Learning Processes and Prior Knowledge. Boston: Kluwer Academic Publishers.

Brown, F. (1978). Principles of Educational and Psychological Testing (2 nd ed.).ILL.: Charles Thomas.

Buros, O. (1977). Fifty years in Testing. Educational Researcher 6 (7) 9 - 15.

Buros, O. (1978). The Eighth Mental Measurement Yearbook. Highland park, N.J.: The Gyphon Press.

Chase, C. (1978). Measurenent for Educational Evaluation (2 nd ed.). Mass.: Addison Wesley

Crocker, L. & Alginas, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart & Winston.

Cronbach, L. (1990). Educational Psychology (5th ed.). New York: Harpex and Row.

Daron, I. & Wolf, D. (Eds.) (1996). Performance - Based Student Assessment. Chicago: National Society for the Study of Education.

Desforges C. (1989). Testing and Assessment .London.: Cassell.

Donald L. & Ernest, L. (1975). Measurement and Evaluation in the Schools. N.J.: Houghton Mifflin Company.

Dubois, P. (1970). A History of Psychological Testing. Boston: Allyn and Bacon.

Dunnette, M. (1966). Personal Selection and Placement. Calif.: Brooks Cole-Publishing Company.

Ebel, R. (1965). Measuring Educational Achievement. Englewood Cliffs, N.J.: Prentice Hall.

Ebel R. (1980). Practical Problems in Educational Measurement. Lexington, MA: D.C. Hearth.

Ebel, R. & Frisbie D. (1991). Essentials of Educational Measurement (5 th ed.).Englewood Cliffs. N. J.: Prentice Hall.

Edwards, A. (1959). Edwards Personal Preference Rocord. San Antonio, Psychological Company.

Elliot, C. (1990). Differential Ability Scales: Introductory and Technical Handbook. San Antonio, TX: Psychological Corporation.

Embertson, S. (1985). Test Design: Development in Psychology and Psychometrics. M.Y. :Academic Press.

Embertson, S. & Reise (2000). Item Response Theory for Proychologists. N.J.: Lawrence .Erlbaum.

Eysenck, H. (1973). The Measurement of Intelligence. Lancaster Medical and Technical Publishing Company.

Eysenck, H. (1979). The Structure and Measurement of Intelligence. N.Y.: Springes - Verlagy.

Farm R. & Tone B. (1994). Portfolios Performance Assessment. N.Y.: Harcourt.

Feldt, L. & Forsyth, R. (1993). Iowa Tests of Educational Development. Iowa City: University of Iowa.

Fredriksen N; Miselevy R. & Bejar I. (1993). Test Theory for a New Generation of Tests.S, N.J., Erlbaum.

Gekoski, N. (1964). Psychological Testing: Theory, Interperation and Practices. Ill.: Charles Thomas Publishing.

Ghiselli, E.(1960). Theory of Psychological Measurement. N.Y.: McGraw - Hill.

Gipps, C. (1994). Beyond Testing: Toward a Theory of Educational Assessment. London: Falmes.

Goldstein, H. & Lewis, T. (eds) (1996). Assessment: Problems, Development and Statistical Issues. Chichester: John Wiley.

Gougy, H. (1953). What Determines the Achievement of High School Students? Journal of Educational Research, 46, 321- 331.

Green, D. (1974). The Aptitude Achievement Distinction., CA.: CTB McGraw Hill.

Gronlund, N. (1968). Readings in Measurement and Evaluation. N. Y.: the Macmillan Company.

Gronluned, N. (1985). Measurement and Evaluation in Teaching. New York: the Macmillan Company.

Gruijter, D. & Kamp, L. (1976). Advances in Psychological and Educational Measurement. N.Y.: Wiley & Sons.

Guilford, J. (1954). Psychometric Methods. N.Y.: McGraw Hill.

Hambleton, R. & Eigner, D. (1978). Guidelines for Evaluating Criterion Referenced Tests and Test Manuals. **Journal of Educational Measurement**, 15 , 321 - 327.

Hambleton, R. & Swaminathan, H. (1985) Item Response Theory: Principles and Applications. Nowell, MA: Kluwer Academic Press.

Hambleton, R. & Zaal, J. (1991). Advances in Educational and Psychological Testing: Theory and Applications. MA: Kluwer Academic Publishers.

Haney, W. (1984). Testing Reasoning and Reasoning About Testing. **Reveiw of Educational Reseach** 54, 4, 597-654.

Harris, D. & Bell, C. (1990). Evaluating and Assessing for Learning. london: kogem Page.

Holmen, M. & Doetor, R. (1972). Educational and Psychological Testing: A Study of the Industry and its Practices.N.Y.: Russell Sage.

Hopkins, K., Stanley, J. & Horbins, B. (1990). Educational and Psychological Measurement and Evaluation (7 th ed.)., Needram Heights, MA: Allyn Bacon.

Horrocks, H. & Schoonover, T. (1968). Measurement for Teachers. Columbus: Charles Merill Publishing Company.

Houts, P. (1977). The Myth of Measurability. N.Y.: Hort Publishing Company.

Hunt, J. (1960). Intelligence and Experience N.Y.: The Ronald Press Company Ingenkamp, K. (1968). Developments in Educational Testing(Volume 1). N.Y.: Gordon and Breach.

Innes, T. (1972) The Prediction of Achievement : Measurement of Schools from Non - School Factors. Invited Address at Southern Invitational Conference on Testing. Athens, Ga.

Jensen, A. (1980). Bias in Mental Testing. N.Y.: Free Press.

Jensen, M. & Feuerstein, R. (1986). Cultural Differences and Cultural Deprivation: A Theoretical Framework for Differential Interventions. In R. M. Goupta & P. Coxhead (eds.), Cultural Diversity and Learning Efficiency. London: the Macmillan Press.

Joint Committee on Testing Practices (1988). Code of Fair Testing Practice in Education. Washington, D.C.: American Psychological Assocation.

Keats, J. (1971). An Introduction to Quantitative Psychology. N.Y.: John Wiley.

Kelly, E. (1967). Assessment of Human Characteristics. Calif.: Brooks Cole Publishing Company.

Kenneth, S.; Melissa, S.; Anderson, R. (2002). The Educational Guide to Ethical Standrads. Washington, D.C.: AERA Publications.

kohn, A. (2000). A case Against Standardized Testing. Portsmouth, NH: Heineman.

koretz, D. (1997). What Happens to Test Scores and Why. **Educational Leadership**, 52 (2) 38. - 41.

Lien, A. (1971). Measurement and Evaluation of Learning. (2nd ed.). Iowa: Brown Company Publishers.

Lindzey, G. (1959). On the Classification of Projective Techniques. Psycholgical Bulletin, 59, 2, 158 - 168.

Linn, R. (1989). Educational Measurement. (3rd ed.) N.Y.: Macmillan.

Linn, R. & Gronlund, N. (1995). Measurement and Evaluation in Teaching (7 th ed.) N.Y.: Merill.

Lohman, D. (1993). Teaching and Testing to Develop Fluid Abilities. **Educational Researcher**, 22 (7) 12 - 23.

Lubin, B.; Larsen, R. & Matarazzo, (1984). Patterns of psychological Test Usage in the United States: 1935 - 1982. **American Psychologist**, 39 , 451 - 454.

Maeroff, G. (1991). Assessing Alternative Assessment Phi Delta Kappan, 72, 272 - 281 .

Marshall, J. & Hales, L. (1971). Classroom Test Construction. Calif.: Addison - Wesley Publishing Company.

Mathematical Sciences Educational Board (MSEB) (1991). A summary of the Regional Presummit Forum. Washington, D.C.: National Research Council.

McCallum, R. (1990). Determining the Factor Structure of The Stanford - Binet. Fourth Edition. **Journal of Psycholeducational Assessment**, 8,436 - 442.

Mclaughlin, M. (1991). Test - Based Accountability as Reform Strategy Kappan 73 (3) 248 - 251.

Mcmillan, J. (1997). Classroom Assessment: Principles and Practice for Effective Instruction. Boston: Allyn and Bacon.

Mehrens, W. & Ebel, R. (1967). Prinicples of Educational and Psychological Measurement Ill.: Rand Mcnally & Company.

Mehrens, /W. & Lehmann, I. (1969). Standardized Tests In Education N.Y.: Holt, Rinehart and Winston, Inc.

Miller, K. (1975). Psychological Testing in Personnal Assessment.: N.Y. John Wiley & Sons.

Millman, I.; Bishop, C. & Ebel, R. (1965). An Analysis of Test Wiseness. **Educational and Psychological Measurement** 25, 707 - 726.

Nitko, J. (1983). Educational Tests and Measurement An Introduction. N.Y.: Harcourt Brace Jovanovich, Inc.

Nitko, J. (1996). Educational Assessment of Students. Englewood Cliffs: Prentice Hall.

Noll, V. & Scannell, D. (1972). Introduction to Educational Measurement (3rd ed.) N.K: Houghton Mifflin Company.

Nunnally, J. (1972). Educational Measurement and Evaluation. N.Y.: MGraw-Hill.

Oosterhof, A. (1990). Clasoroom Applications of Educational Measurement. Columbus: Merill.

Payne, D. (1974). The Assessment of Learning: Cognitive and Affective. Mass.: D.C. Heath and Company.

Perin L. (1975). Personality Theory Assessment and Research. N.Y.: John Wiley.

Phye, G.(1996). Handbook of Classroom Assessment: Learning, Adjustment, and Achievement N.Y. Academic press.

Popham, W. (1978). Crieterion- Referenced Measurement. N.J. Prentice Hall, Inc.

Popham, W. (1981). Modern Educational Measurement Educational (Engel-wood Cliffs). N.T:

Popham, W. (1995). Classroom Assessment: What Teachers Need to Know. Boston: Allyn And Bacon.

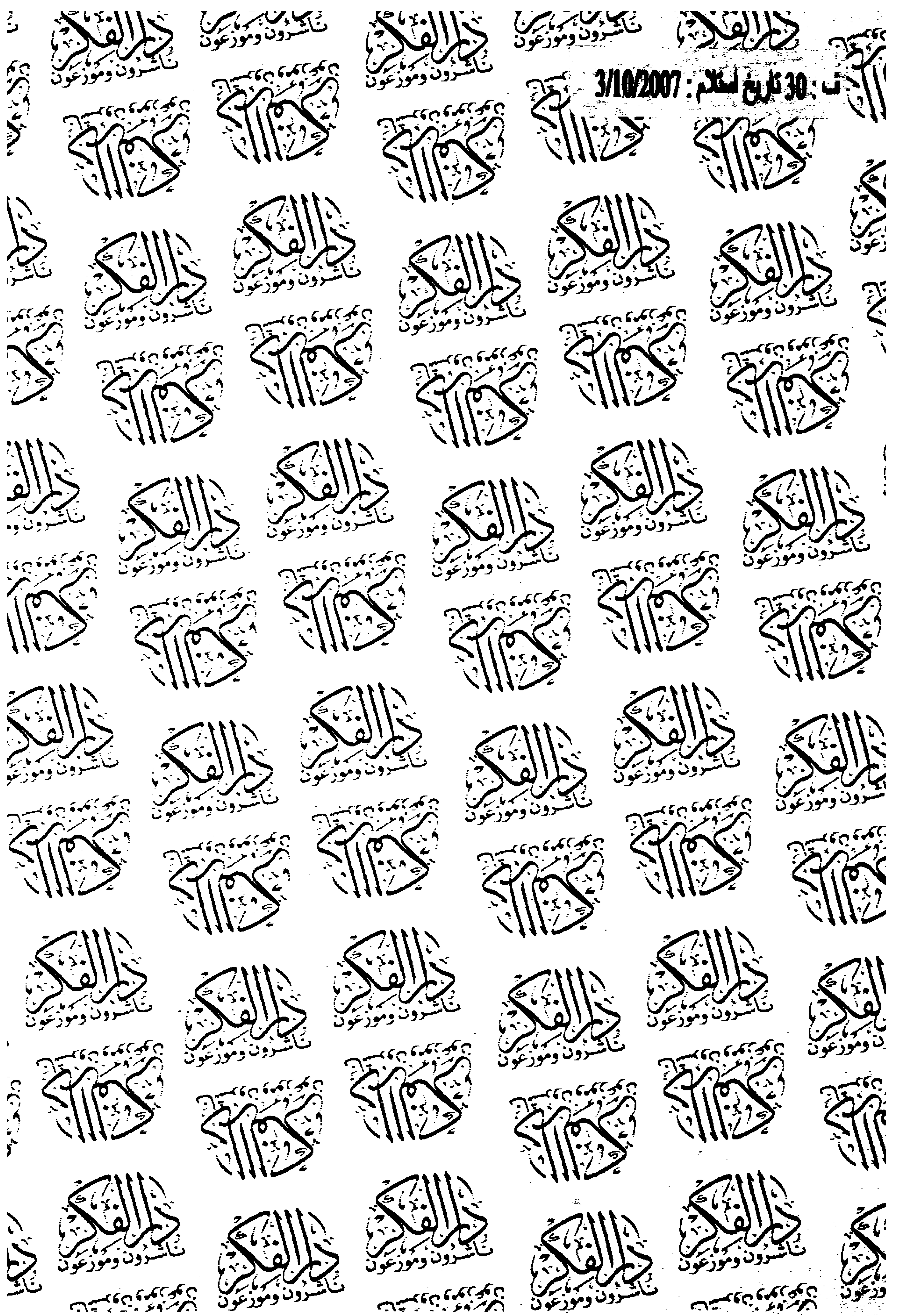
Psychological Corporation (1993). Metropoliatan Achievement Tests. San Antonio Tx:, Author.

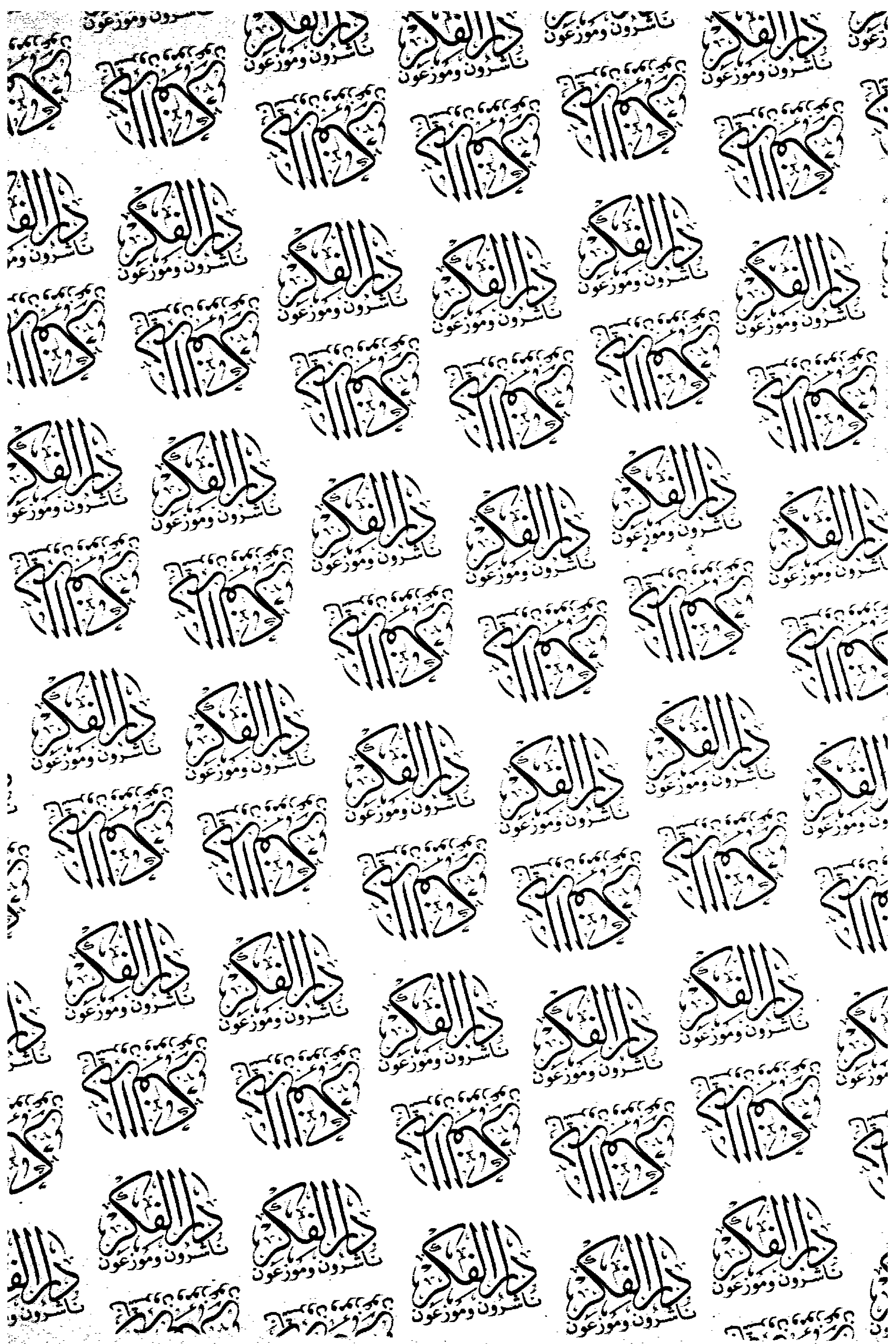
Rabin, A. (1968). Projective Techniques in Personality Assessment: Modern Introduction. N.Y.: Springes Publishing Company, Inc.

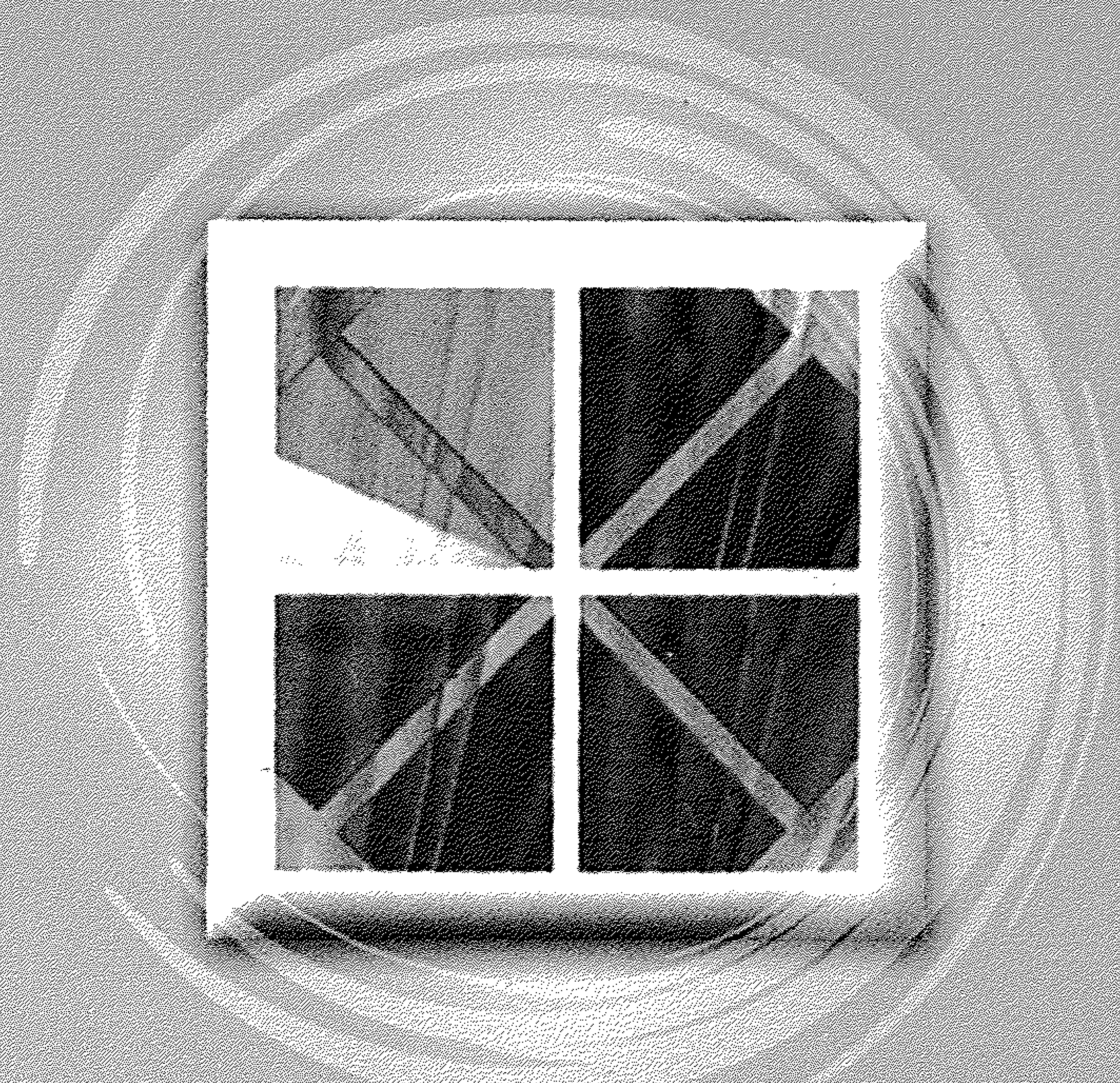
Reschly, D. & Wilson, M. (1990). Cognitive Processing Versus Traditional Intelligance. **School Psychology Review**,191,443-458.

Resnick, L. (1976). The Nature of Intelligence. N.J.: Lawrence Erlbaum.

Roaed, G. (1989). Item Writing and Item Banking by Microcomputers An: Update. **Educational Measurement: Issues and Practice**. 8 (3) 38, 17 - 20.


دار الفكر
ناشرون وموزعون


دار الفكر
ناشرون وموزعون



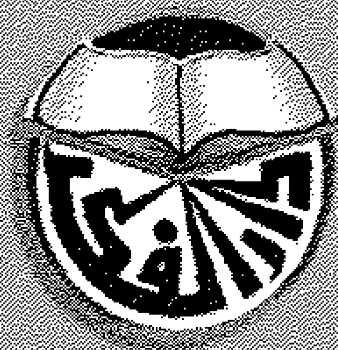
الإختبارات والمقاييس التربوية والنفسية

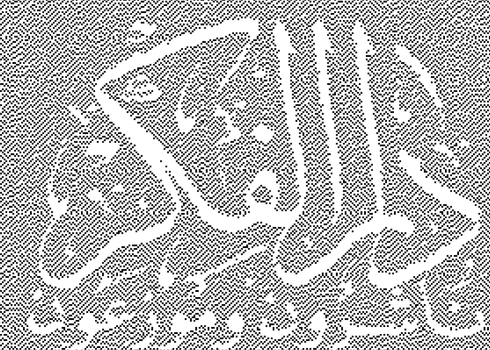




ISBN 9957-07-437-7
9 789957 074371







عمان . ساحة الجامع الحسيني . سوق البتراء . عمارة الحجيري
هاتف 4621938 فاكس 4654761 ص.ب. 183520 عمان 11118 الأردن
www.daralfiker.com info@daralfiker.com